AN INTEGRATED SYSTEMS BIOLOGY APPROACH TO BETTER UNDERSTAND CANCER

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

David (Wei) Dai

August 2019

AN INTEGRATED SYSTEMS BIOLOGY APPROACH TO BETTER UNDERSTAND CANCER

David (Wei) Dai, Ph.D.

Cornell University 2019

Cancer is an umbrella term that encompasses a collection of related diseases. In all types of cancer, as cells become unregulated, they begin to divide uncontrollably and spread into surrounding tissues. From early to late stages of disease progression, there are many alterations that occur and many criteria that must be met for cancer to proliferate uncontrollable, resist cell death, avoid the immune system, and metastasize. Targeted therapy have been developed for a wide range of dysregulated cancer pathways using different modalities and mechanism of actions. However, due to the diversity of cancer, the same treatment that may be effective for one type of cancer may not be responsive on another. Furthermore, patients may carry resistance forms of the disease. To alleviate this, new strategies have been developed to target specific mutations of known dysregulated proteins and to use combination therapies that target multiple pathways. However, both methods require vast amount of knowledge on the biological interactions and mechanism of actions that takes place within the cell. To address this knowledge gap, we believe that the metabolism can be used as a tool to better understand the dysregulations of signaling and gene expression. There is a great opportunity to study the system as a whole to gain key insights for combination therapies that target different regulatory pathways, such as the metabolism and signaling. In this work, we leveraged our current understanding of signaling transduction, transcription factor, and

metabolic networks to develop an integrated systems biology approach to quantitatively unravel the mechanisms that regulate cancer. Ultimately, we were able to establish a computational model that incorporated mechanistic understanding of multiple layers of cellular decision making. We believe this work will be useful in the development and evaluation of new combination therapy across all form of cancers.

BIOGRAPHICAL SKETCH

Born in Chongqing China, David was the eldest son of Zhengyuan Dai, a university professor, and Li Meng, an accountant. David spent his early childhood exploring the campus of Chongqing University where his dad was a faculty in the department of mechanical engineering. On February 8th, 2000, in pursuit of opportunity and a bright future for their song, the Zhengyuan and Li left behind their comfortable life for a one way ticket to Toronto, Canada. Wanting to follow his dad's footsteps in engineering, David obtained a Bachelors of Science in Chemical Engineering from University of Maryland Baltimore County in the May of 2014. In the June of 2014, he joined Professor Jeffrey Varner's lab at Cornell University taking on a project to develop a novel framework to simulate constraint-based metabolic that can take account into intracellular spatial variation and diffusion limitations. Towards the end of that project, David gained interest in the applications of metabolic engineering and systems biology in human health and pursued a collaboration to better understand cancer metabolism and it's role in tumor progression with Professor Stroock. David spent the next 4 years at Cornell developing computational frameworks to understand the regulations of cancer metabolism. Received his doctorate in chemical engineering in the Summer of 2019, and moved on to a senior scientist position at Pfizer supporting the oncology research unit with quantitative systems pharmacology modeling.

This work is dedicated to my mom and dad.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude towards my thesis advisor Professor Jeffrey Varner for his guidance, patience, enthusiasm, and his obsession for the Helvetica font and Apple products. Jeff, not only did you teach me how to solve problems, more importantly you taught me the importance of communicating and selling your ideas. The most important lesson I learned at Cornell was not how to conduct research, but rather what I need to do to convey those ideas and why they are important, the big picture stuff. Thank you. I would also like to thank Professors Abraham Stroock for the endless career and life advice, and guidance. Thank you for all the discussions when I was lost about the path after graduate school. Thank you Professor Matthew Paszek, for the research discussions and an important lesson on integrity. My sincere gratitude also goes out to my undergraduate professors at UMBC who inspired me to pursue graduate studies: Professor Jennie Leach, Professor Maria Jose Castellanos, Professor Taryn Bayles, and Professor Joseph Loehe. This work could not have happened without the mentorship of the members of the Yen lab: Andy Yen, Petruta Rodica, and Rob McDonald. Thank you Mengrou Shan for being such an amazing friend and collaborator. Thank you wine friends and volleyball friends for keeping me sane and challenging me outside of academics. Lily Johnson, thank you for being a supportive friend and always making time to read over my work and providing insightful feedback. You really shaped the way I write today. Kathy Rogers and Holly Jensen, thank you so much for all thee long email replies and career advices you have shared with me. Mason Minot, Adi Sagar, and Desiigner, this work would be incomplete without you. Thank you for always checking up on me and having my back even after all these years. Jasmin Kennard, thank you for always being there for me. You are

a friend I can always rely on and trust. Alex Chen, thank you for being you. Nike Horvath, thank you showing me the jovial side of life, my experience at Cornell would not have been the same without you. Our shenanigans was therapeutic for my mental health. Mike Vilkhovoy, thank you for always having my back and being such an amazing friend that I can count on. You always pushed me to do better, words don't come close to my gratitude for the last 5 years with you. Thank you Chelsea Hu for being the older sister that I never wanted, but really needed. Thank you for looking out for me in the last 5 years. Finally, I would like to thank my parents for their unconditional love, undying support, endless sacrifices. Without your lessons on life, I would not be here today.

TABLE OF CONTENTS

Biographical Sketch . . . iii
Dedication . . . iv
Acknowledgements . . . v
Table of Contents . . . vii
List of Tables . . . ix
List of Figures . . . x

1 Introduction **1**
1.1 Mathematical Modeling of Biological Networks . . . 1
1.1.1 Kinetic Modeling . . . 1
1.1.2 Constraint-based Modeling . . . 5
1.2 Current Approach in Cancer Therapy . . . 11
1.3 Cancer Metabolism . . . 13
1.4 HL60 differentiation . . . 15
1.5 Outlook . . . 17

2 An Effective Model of the Retinoic Acid Induced HL-60 Differentiation Program **19**
2.1 Abstract . . . 19
2.2 Introduction . . . 20
2.3 Results . . . 23
2.4 Discussion . . . 32
2.5 Methods . . . 39

3 Modeling and Analysis of Combination Treatment of HL60 Transcription Factor Network **65**
3.1 Abstract . . . 65
3.2 Introduction . . . 66
3.3 Results . . . 69
3.4 Discussion . . . 74
3.5 Materials and Methods . . . 77

4 Src family kinase inhibitor bosutinib enhances retinoic acid-induced differentiation of HL-60 leukemia cells **95**
95
4.2 Introduction . . . 96
4.3 Materials and Methods . . . 98
4.4 Results . . . 100
4.5 Discussion . . . 102

5 Dynamic Sequence Specific Constraint-based Modeling of Cell-free Protein Synthesis **111**
111
5.2 Introduction . . . 112
5.3 Results . . . 117
5.3.1 Cell-free *E. coli* metabolic network. . . . 117
5.3.2 Dynamic constrained simulation of cell free protein synthesis. . . . 118
5.3.3 Alternative measurement sets. . . . 121
5.4 Discussion . . . 125
5.5 Materials and Methods . . . 130
5.5.1 Formulation of the model equations. . . . 130
5.5.2 Generation and evaluation of alternative measurement sets. 134

6 Metabolic characterization of ATRA-induced HL60 differentiation **161**

7 On going and Future Works **170**
7.1 Response Resistance and Rescue - Metabolic characterization of ATRA-induced differentiation . . . 170
7.2 Connecting Signaling to Metabolism using Kinetic and Constraint-based Modeling . . . 172
7.3 Expansion of Framework into Triple Negative Breast Cancer . . . 176

8 Concluding Remarks **179**

A Models Made Available **180**

Bibliography **181**

LIST OF TABLES

2.1 Myelomonocytic transcription factor connectivity used in the signal integration and phenotype modules. 60
2.2 Characteristic model parameters estimated from literature. . . . 62
2.3 Sequence lengths from NCBI RefSeq database were used in the signal integration and phenotype modules [184]. The RNA sequence length used represents the total distance of transcription, and assume to be equal to the gene length. 64

3.1 Myelomonocytic transcription factor connectivity used in the signal integration and phenotype modules. 93

5.1 Reference values for transcription, translation, and mRNA degradation from literature. Transcription rate calculated from elongation rate, mRNA length, and promoter activity level. Translation rate calculated from elongation rate, protein length, and polysome amplification constant. The mRNA degradation rate calculated from a characteristic mRNA half-life. 139
5.2 Flux uncertainty calculated using flux variability analysis for the base synthetic dataset during the first production phase (0 h to 1.5 h), normalized to the glucose consumption rate. 140
5.3 Flux uncertainty calculated using flux variability analysis for the base synthetic dataset during the second production phase (1 h to 3 h), normalized to the glucose consumption rate. 146
5.4 Metabolites by frequency of appearance in the simulated annealing constraint sets. 152
5.5 Metabolites by frequency of appearance in the 57 best simulated annealing constraint sets, those with error at least three orders of magnitude lower than the base synthetic dataset. 153

LIST OF FIGURES

1.1 Schematic of how a system of ordinary differential equations (ODEs) are implemented to represent a network of biological interactions. The rate of change with respect to time of each individual species is represented by a mass balances. The rate of accumulation or depletion of the species are determined by the rate of reactions. The diagram shows an example of a substrate (S) being converted to product (P) at the rate of v_1, while the presence of drug (D) inhibits the reaction. 3

1.2 Simplified schematic of the human complement system. The complement cascade is activated through three pathways: the classical, the lectin, and the alternative pathways. Complement initiation results in the formation of classical or alternative C3 convertases, which amplify the initial complement response and signal to the adaptive immune system by cleaving C3 into C3a and C3b. C3 convertases further react to form C5 convertases which catalyze the cleavage of the C5 complement protein to C5a and C5b. C5b is critical to the formation of the membrane attack complex (MAC), while C5a recruits an adaptive immune response. 4

1.3 Flux balance analysis is formulated as a linear programming problem with an objective function to be maximized or minimized subjected to a set of constraints on the fluxes and the rate of accumulation of depletion. 8

1.4 (Caption next page.) 9

1.4 (Previous page.) Multi-scale modeling of cancer metabolism [204]. (A) Flux Balance Analysis (FBA). The arrows represent fluxes of species within a reduced representation of cell metabolism and cell growth. Key steps associated with three hypotheses are labeled: (1) Warburg effect, (2) Reverse Warburg and (3) Glutamine addiction. The uptake and production rates are qi/n [g/g-DW-hr] for the ith metabolite and the nth metabolic phenotype. We impose the maximum growth rate, ,n [hr-1], for a given metabolic phenotype of each cell type as an objective function within the FBA. (B) Cell: Biomass (Xm [g]) growth of each cell type is modeled as a Monod-like process parameterized by the same maximum growth rates used in FBA that are modulated by functions of metabolite concentrations, fn(Cj)Monod. The change in Volume of the cell (Vm [L]) is calculated from biomass growth by applying a constant density of the cell ([g-DW/L]). Yield coefficients (Yi/n [g-DW/g]) for each metabolite (i) and corresponding metabolic phenotype (n) are defined in terms of the uptake and production rates (qi/n) obtained from FBA. Extracellular space: Species balances for each explicit metabolite follow reaction-diffusion kinetics and govern the concentration profiles of metabolite at the multicellular scale. These equations (Eqs 1-4 in text) are integrated into and solved within an agent-based model (ABM; iDynoMiCS). (C) ABM simulations: i) Radial, two dimensional growth: Tumor cells grow radially out from an initial cluster of cells with metabolites supplied at the edge of the cell mass such that radial gradients of concentration emerge (color map;red is high and blue is low concentration). Two phenotypes are displayed (red: tumor cells and blue: stromal cells). As the tumor grows, concentration gradients of metabolites become significant, making the tumor growth a diffusion-limited process that can result in different growth dynamics as well as distinct spatial distribution of cell subpopulations. ii) Axial, one dimensional growth: Layers of tumor cells (red) and stromal cells (blue) are initiated near a blood vessel that supplies metabolites (from the top), such as glucose and oxygen in the blood stream. Growth pushes cells deeper into the tissue, away from the vessel, such that strong gradients of metabolite can again occur. iii) Krogh length calculation: To evaluate the impact of diffusion limitations in a simple model, we treat cells as continuum with uniform, zeroth order kinetics of metabolite consumption to calculate the distance over which the concentration of limiting metabolites falls to zero within the tumor mass; we refer to this distance as the Krogh length of a given metabolite. 10

2.1 Schematic of the effective ATRA differentiation circuit. Above a critical threshold, ATRA activates an upstream Trigger, which induces signalsome complex formation. Signalsome activates the mitogen-activated protein kinase (MAPK) cascade which in turn drives the differentiation program and signalsome formation. Both Trigger and activated cRaf-pS621 drive a phenotype gene expression program responsible for differentiation. Trigger activates the expression of a series of transcription factors which in combination with cRaf-pS621 result in phenotypic change. . . 51
2.2 Model analysis for ATRA-induced HL-60 differentiation. A: BLR1 mRNA versus time following exposure to 1μM ATRA at t = 0 hr. B: cRaf-pS621 versus time following exposure to 1μM ATRA at t = 0 hr. Points denote experimental measurements, solid lines denote the mean model performance. Shaded regions denote the 99% confidence interval calculated over the parameter ensemble. C: Signalsome and cRaf-pS621 nullclines for ATRA below the critical threshold. The model had two stable steady states and a single unstable state in this regime. D: Signalsome and cRaf-pS621 nullclines for ATRA above the critical threshold. In this regime, the model had only a single stable steady state. E: Morphology of HL-60 as a function of ATRA concentration (t = 72 hr). Experimental data in panels A and B were reproduced from Wang and Yen [242], data in panel E is reported in this study. 52
2.3 Model simulation following exposure to 1μM ATRA. A: BLR1 mRNA versus time with and without MAPK inhibitor. B: cRaf-pS621 versus time following pulsed exposure to 1μM ATRA with and without BLR1. Solid lines denote the mean model performance, while shaded regions denote the 99% confidence interval calculated over the parameter ensemble. C: Western blot analysis of phosphorylated ERK1/2 in ATRA washout experiments. Experimental data in panels A and B were reproduced from Wang and Yen [242], data in panel C is reported in this study. The images of the raw gel for panel C is given in the Supplemental Materials. 53

2.4 Model simulation of the HL-60 gene expression program following exposure to 1μM ATRA at t = 0 hr. A: Scaled CD38 and CD11b expression versus time following ATRA exposure at time t = 0 hr. B: Scaled Gene expression at t = 48 hr following ATRA exposure. Gene expression was normalized to expression in the absence of ATRA. The gene expression is quantified by the protein fold change of quantified Western blot data (from at least three biological repeat nuclear lysates) using ImageJ. Experimental data in panels A and B were reproduced from Jensen et al. [119]. Model simulations were conducted using the ten best parameter sets collected during model identification. Solid lines (or bars) denote the mean model performance, while the shaded region (or error bars) denote the 95% confidence interval calculated over the parameter ensemble. 54
2.5 Model simulation of HL-60 cell-cycle arrest following exposure to 1μM ATRA at t = 0 hr. A: Predicted p21 and E2F expression levels for the best parameter set following ATRA exposure at time t = 0 hr. B: Estimated fraction of HL-60 cells in G0 arrest following ATRA exposure at time t = 0 hr. Solid lines (or bars) denote the mean model performance, while the shaded region (or error bars) denotes the 95% confidence estimate of the polynomial model. Experimental data in panel B was reproduced from Jensen et al. [119]. 55
2.6 Robustness of the HL-60 differentiation program following exposure to 1μM ATRA at t = 0 hr. A: Singular value decomposition of the average system response (l^2-norm between the perturbed and nominal state) following pairwise gene knockout simulations using the top ten best fit parameter sets. The rows denote the deleted genes, while columns denote the response mode. B: Singular value decomposition of the average system response (l^2-norm between the perturbed and nominal state) following the pairwise removal of protein-DNA connections for the top ten best fit parameter sets. The rows denote protein-DNA interactions at the labeled promoter, while the columns denote the top ranked response modes (combinations of deletions). The percentage at the top of each column describes the fraction of the variance in the system state captured by the node combinations in the rows. 56
2.7 Robustness of the HL-60 differentiation program following exposure to 1μM ATRA at t = 0 hr. Protein fold change at t = 48 hr (rows) in single and double knock-out mutants (columns) relative to wild-type HL-60 cells. The responses were grouped into >2,4 and 8 fold changes. The best fit parameter set was used to calculate the protein fold change. 57

2.8 Investigation of a panel of possible Raf interaction partners in the presence and absence of ATRA. A: Species identified to precipitate out with Raf: first column shows Western blot analysis on total Raf immunoprecipitation with and without 24 hr ATRA treatment and the second on total lysate. B: The expression of species considered that did not precipitate out with Raf at levels detectable by Western blot analysis on total lysate. C: Effect of the Raf inhibitor GW5074 on Raf interactions as determined by Western blot analysis of total Raf immunoprecipitation. The Authors note the the signal associated with Src was weak. D: Cell Cycle distribution as determined by flow cytometry indicated arrest induced by ATRA, which was increased by the addition of GW5074. E: Expression of the cell surface marker CD11b as determined by flow cytometry indicated increased expression induced by ATRA, which was enhanced by the addition of GW5074. F: Inducible reactive oxygen species (ROS) as determined by DCF flow cytometry. The functional differentiation response of ATRA treated cells was mitigated by GW5074. GAPDH was used as a loading control. The black lines frame groupings from independent gels and each image is typical of three repeats. Images of the raw gels for panels A, B C and F are given in the Supplemental Materials. 58

2.9 This schematic diagram shows the hypothetical principal pathways in the ATRA–induced signaling that results in cell differentiation in the HL-60 myeloid leukemia model [238, 34, 35, 208, 21, 207]. It is based on modules and feedback loops. There are three main arms (shown top to bottom): 1. Direct ATRA targeting of RAREs in genes such as CD38 or BLR1; 2. Formation of a signalsome that has a regulatory module that includes Vav (a guanine nucleotide exchange factor), CBL and SLP-76 (adaptors), and Lyn (a Src family kinase) that regulates a Raf/Mek/Erk axis that incorporates Erk to Raf feedback, where the regulators are modulated by AhR and CD38 receptors; and 3. Direct ATRA targeted up regulation of CDKI to control RB hypophosphorylation. The Raf/Mek/Erk axis is embedded in the signalsome and subject to modulation by the regulators. The output of the signalsome is discharge of the Raf from the cytosol to the nucleus where it binds (hyper)phospho-RB and other targets, including NFATc3, which enables activation of the ATRA bound RAR/RXR poised on the BLR1 promoter, and also GSK3, phosphorylation of which relieves its inhibitory effect on RARα. CDKI directed hypophosphorylation of RB releases Raf sequestered by RB to go to NFATc3, GSK3, and other targets. A significant consequence of the nuclear RAF is ergo ultimately to enable or hyperactivate transcriptional activation by RARα to drive differentiation. It might be noted that this proposed general model provides a mechanistic rationalization for why cell cycle arrest is historically oft times perceived as a precondition for phenotypic maturation. 59

3.1 HL60 transcription factor network structure from literature sources. Figure shows interactions between protein nodes with black arrows indicating transcriptional activation and red lines indicating transcriptional repression. Phenotypic markers are shown on the right, with black indicating activation and red repression. 85

3.2 ATRA, D3, and combination induced expression of key proteins in the gene regulatory network at 48 hours after induction. The error bars represent standard errors, 491 samples for simulated data, and 3 samples for experimental measured data. The dotted horizontal line indicated the baseline level of protein at time 0hr. 86

3.3 Dynamic phenotypic markers of ATRA, D3, and combination induced differentiation of HL60. The simulation captured time-course experimental measurements of CD38 (blue), CD11b (red), and CD14 (purple). The points are experimental measurements with error bars representing standard error. The black lines are the simulated average of the 491 parameter set ensemble, with the darker colored band representing 99 % standard error of the simulated mean, and the lighter colored bands meaning 99 % confidence interval. 87
3.4 Sensitivity analysis comparing ATRA and D3 induced HL60 differentiation. The parameters are ranked ordered from low to high sensitivity with the parameters that are more sensitive for ATRA and D3 treatment fall on the right and left side of the 45 degree line respectively. The diamonds represent the weight parameters (W), the squares are saturation constants (K_s), and the circles are cooperativity coefficients (n). 88
3.5 Sensitivity analysis comparing ATRA and D3 induced HL60 differentiation. The parameters are ranked ordered from low to high sensitivity with the parameters that are more sensitive for ATRA and D3 treatment fall on the right and left side of the 45 degree line respectively. The diamonds represent the weight parameters (W), the squares are saturation constants (K_s), and the circles are cooperativity coefficients (n). 89
3.6 Sensitivity analysis comparing D3 and combination induced HL60 differentiation. The parameters are ranked ordered from low to high sensitivity with the parameters that are more sensitive for D3 and combination treatment fall on the right and left side of the 45 degree line respectively. The diamonds represent the weight parameters (W), the squares are saturation constants (K_s), and the circles are cooperativity coefficients (n). 90
3.7 The frequency of off-axis appearance of ranked-order sensitivity comparison between different treatments. The grey circle indicate ATRA vs D3 comparison, with ATRA sensitive parameters in the positive region and D3 sensitive parameters in the negative. The blue data is the ATRA vs combination comparison, while the red is the D3 vs combination comparison. 91
3.8 Log of the 36 cost functions associated all three treatment plotted for three different 495-parameter-set-ensemble. Blue is the ensemble generated using parameter estimation, red and purple are ensembles generated randomly and using Sobol sequences respectively. 92

4.1 Phenotypic analysis of HL-60 cells treated with RA/bosutinib. (A) HL-60 cells were cultured in the presence of 1RA or 1RA and 0.25bosutinib as indicated. CD38 expression was assessed by flow cytometry following 24, 48, and 72h treatment periods. Gating to discriminate positive cells was set to exclude 95% of untreated controls (n=3). Error bars indicate SEM. (B) Normalized means of CD38 expression per cell at 24, 48, or 72h (n=3). (C) CD11b expression was assessed by flow cytometry at 48 and 72h (n=3). *p¡.05 comparing RA-treated samples to RA/bosutinib-treated samples. Two-tailed paired-sample t-tests were used to determine significance. (D) Normalized means of CD11b expression per cell at 48and 72h (n=3). *p¡.05 comparing RA-treated samples to RA/bosutinib-treated samples. (E) Cell cycle distribution showing the percentage of cells in G1/G0 was analyzed using flow cytometry with propidium iodide staining at 24, 48, and 72h (n=4). *p¡.05 comparing RA-treated samples to RA/bosutinib-treated samples. (F) Cell counts were taken at 24, 48, and 72h using a hemocytometer and 0.2% Trypan Blue exclusion staining (n=3). 106

4.2 HL-60 cells treated with RA/bosutinib displayed enhanced respiratory burst and p47phox expression. (A) HL-60 cells were cultured in the presence of 1RA or 1RA and 0.25bosutinib as indicated. Respiratory burst was analyzed by measuring inducible reactive oxygen species (ROS) production by flow cytometry using the 2,7-dichlorofluorescein (DCF) assay. Gates to determine percent increase of expression with treatment were set to exclude 95% of the DMSO-treated control population for each culture condition; TPA-treated samples show induced ROS (n=3). Error bars indicate SEM. *p¡.05 comparing RA-treated samples to RA/bosutinib-treated samples. Two-tailed paired-sample t-tests were used to determine significance. (B) HL-60 cells were cultured for 48h in the presence of 1RA or 1RA and 0.25B as indicated and whole cell lysate was collected. Twenty five microgram of lysate per lane was run. Western blots of PAGE-resolved lysates were probed for p47phox (n=3). Films were scanned and bands of interest were quantified using NIH ImageJ. Error bars indicate SEM. A representative blot, cropped to show only the band of interest, is included. (C) Western blots of GAPDH were used as loading controls following the procedure described above. 107

4.3 Bosutinib enhances RA-induced SFK expression and diminishes SFK phosphorylation. (A) HL-60 cells were cultured for 48h in the presence of 1RA or 1RA and 0.25bosutinib (B) as indicated and whole cell lysate was collected. Twenty five microgram of lysate per lane was run. Western blots of PAGE-resolved lysates were probed for Fgr (n=3). Films were scanned and bands of interest were quantified using NIH ImageJ. Error bars indicate SEM. A representative blot, cropped to show only the band of interest, is included. (B) Western blots of Lyn following the procedure described above. (C) Western blots of phosphorylated pan-Y416 SFK following the procedure described above. (D) Western blots of GAPDH were used as loading controls following the procedure described above. 108
4.4 Bosutinib enhances levels of RA-induced phosphorylated c-Raf. (A) HL-60 cells were cultured for 48h in the presence of 1RA or 1RA and 0.25bosutinib as indicated and whole cell lysate was collected. Twenty five microgram of lysate per lane was run. Western blots of PAGE-resolved lysates were probed for c-Raf pS259 (n=3). Films were scanned and bands of interest were quantified using NIH ImageJ. Error bars indicate SEM. A representative blot, cropped to show only the band of interest, is included. (B) Western blots of c-Raf pS621 following the procedure described above. (C) Western blots of c-Raf pC-terminal domain following the procedure described above. (D) Western blots of total c-Raf following the procedure described above. (E) Western blots of GAPDH were used as loading controls following the procedure described above. 109
4.5 The effect of bosutinib on c-Cbl, mTOR, AhR, and p85 PI3K. (A) HL-60 cells were cultured for 48h in the presence of 1RA or 1RA and 0.25bosutinib (B) as indicated and whole cell lysate was collected. Twenty five microgram of lysate per lane was run. Western blots of PAGE-resolved lysates were probed for c-Cbl (n=3). Films were scanned and bands of interest were quantified using NIH ImageJ. Error bars indicate SEM. A representative blot, cropped to show only the band of interest, is included. (B) Western blots of mTOR following the procedure described above. (C) Western blots of AhR following the procedure described above. (D) Western blots of p85 PI3K following the procedure described above. (E) Western blots of GAPDH were used as loading controls following the procedure described above. 110

5.1 Schematic of the core portion of the cell-free *E. coli* metabolic network. The network consisted of 264 reactions and 146 metabolites. Metabolites of glycolysis, pentose phosphate pathway, Entner-Doudoroff pathway, and TCA cycle are shown. Metabolites of oxidative phosphorylation, amino acid biosynthesis and degradation, transcription/translation, chorismate metabolism, and energy metabolism are not shown. 154
5.2 Simulated metabolite concentration versus synthetic data as a function of time. Central carbon metabolism, including glucose (substrate), CAT (product), and intermediates, as well as total concentration of energy species (energy total). The energy total denotes the summation of all energy species in the model (all bases and all phosphate states). The 95% confidence interval for the simulation conducted over the ensemble of transcription/translation parameter sets is shown in the orange shaded region, while the 95% confidence interval for the synthetic constraint data is shown in the blue shaded region. The synthetic data constraints were generated from the kinetic model of Horvath et al, which was trained using experimental measurements of the system simulated in this study [103]. 155
5.3 Simulation of amino acid concentration versus synthetic data as a function of time. The 95% confidence interval for the simulation conducted over the ensemble of transcription/translation parameter sets is shown in the orange shaded region, while the synthetic constraint data is shown in the blue shaded region. Arginine and glutamate were excluded from the constraint set. The synthetic data constraints were generated from the kinetic model of Horvath et al, which was trained using experimental measurements of the system simulated in this study [103]. 156
5.4 Simulation of energy species and energy totals by base versus synthetic data as a function of time. The 95% confidence interval for the simulation conducted over the ensemble of transcription/translation parameter sets is shown in the orange shaded region, while the synthetic constraint data is shown in the blue shaded region. The synthetic data constraints were generated from the kinetic model of Horvath et al, which was trained using experimental measurements of the system simulated in this study [103]. 157
5.5 Flux uncertainty versus metabolite prediction error against synthetic data, normalized to the base case (white star), for exclusion (gray) and inclusion (black) metabolite constraint sets. The performance of the SVD-determined metabolite constraint set is shown by the white square. 158

5.6 Flux uncertainty and metabolite prediction error for the simulated annealing experimental design approach. **A**: Normalized flux uncertainty versus normalized metabolite prediction error. **B**: Number of metabolite constraints versus normalized metabolite prediction error. Error was computed for the synthetic experimental designs normalized to the base synthetic dataset (white star). Sets that include glucose are show as gray circles, while those that do not are represented with black circles. 159

6.1 HL60 culture seeded at 100k cells/mL cultured for 72 hours. Growth curve, glucose and lactate media concentration. The light gray represent untreated cultures while the dark gray are the ATRA-treated cultures. The error bars represent one standard deviation of three biological replicates. The arrows represents the exchange of metabolic fluxes. 163
6.2 HL60 culture seeded at 100k cells/mL cultured for 72 hours. Serine, glycine, and cysteine media concentration. The light gray represent untreated cultures while the dark gray are the ATRA-treated cultures. The error bars represent one standard deviation of three biological replicates. The arrows represents the exchange of metabolic fluxes. 164
6.3 HL60 culture seeded at 100k cells/mL cultured for 72 hours. Alanine, valine, leucine, phenylalanine, and tryptophan media concentration. The light gray represent untreated cultures while the dark gray are the ATRA-treated cultures. The error bars represent one standard deviation of three biological replicates. The arrows represents the exchange of metabolic fluxes. 165
6.4 HL60 culture seeded at 100k cells/mL cultured for 72 hours. Histidine, glutamine, glutamate, and proline. media concentration. The light gray represent untreated cultures while the dark gray are the ATRA-treated cultures. The error bars represent one standard deviation of three biological replicates. 166
6.5 HL60 culture seeded at 100k cells/mL cultured for 72 hours. Arginine, aspartate, threonine, lysine, methionine, isoleucine. media concentration. The light gray represent untreated cultures while the dark gray are the ATRA-treated cultures. The error bars represent one standard deviation of three biological replicates. 167
6.6 Flux balance analysis model with metabolite measurement constraints. Light gray indicates untreated HL60 fluxes while Dark gray indicates ATRA-treated HL60 fluxes. All fluxes are in the units of mmol/gDW. 168
6.7 Schematic approach in connecting the different hierarchy of cellular decision making. 169

7.1 The MDA-MB-231 reporter cell line consists of a transfected vector that contains a NANOG promotor directly upstream of GFP. Therefore in the presence of NANOG, green fluorescent protein (GFP) will be produced, differentiation the high expressing NANOG cells from the low and null in the total population. In this study, we defined the high and low NANOG-expressing cells to be top and bottom 5% of the GFP positive cells, and null to be the subpopulation that are GFP negative. Leveraging our metabolomic efforts, we used the change in metabolite concentrations in the media as constraints for the FBA model to determine intracellular flux estimates for each subpopulations. 178

CHAPTER 1

INTRODUCTION

1.1 Mathematical Modeling of Biological Networks

The use of computational tools and mathematical modeling has long contributed to our understanding of biochemical networks, in particular, there is a long history of quantitative mechanistic modeling [6]. Currently, there are many existing mathematical approaches to characterize biochemical networks such as cellular metabolism and its regulation [26, 71, 51, 8, 152, 180, 138, 186, 36, 247]. The two modeling approaches that will be discussed in this work are the kinetic modeling and constraint-based modeling approaches.

1.1.1 Kinetic Modeling

Kinetic models are constructed by creating a mathematical representation of biological interactions of a network. This approach is one of the most common modeling approaches for signaling transaction networks using mass action kinetics. ODE kinetic models often require extensive prior knowledge of network structure, rate constants and initial conditions [131]. An example of how ODE models can be implemented is shown in Fig 1.1. Using this methodology, Lauffenburger and coworkers developed early biophysical and kinetic models of epidermal growth factor receptor (EGFR) signaling in fibroblastic cells and interleukin 2 receptor signaling in T-cells [78, 214]. Additional ODE models were developed focusing on downstream signaling due to the presence of growth factors and its effect on cell fate decisions [200, 132]. ODE models have also

been used in the studies of enzyme facilitated reaction networks. For instance, the human complement system is a complex system where proteases cleave inactive substrate into active product (Fig.1.2). Interestingly, these active products can combine to produce new proteases that facilitate downstream reactions therefore acting as a species and an enzyme.

Complement is an important pathway in innate immunity. The complement cascade involves many soluble and cell surface proteins, receptors and regulators [239, 240]. Traditionally, complement models have been formulated as systems of linear or non-linear ordinary differential equations (ODEs). For example, Hirayama et al., modeled the classical complement pathway as a system of linear ODEs [95], while Korotaevskiy and co-workers modeled the classical, lectin and alternative pathways as a system of non-linear ODEs [139]. More recently, large mechanistic models of sections of complement have also been proposed. For example, Liu et al., analyzed the formation of the classical and lectin C3 convertases, and the regulatory role of C4BP using a system of 45 non-linear ODEs with 85 parameters [154]. Zewde and co-workers constructed a detailed mechanistic model of the alternative pathway which consisted of 107 ODEs and 74 kinetic parameters and delineated between the fluid, host and pathogen surfaces [259]. Recently, we estimated an ensemble of experimentally validated reduced order complement models using multiobjective optimization [197]. The modeling approach combined ordinary differential equations with logical rules to produce a complement model with a limited number of equations and parameters. Although we will not discuss the model in detail for this thesis, this earlier work of reduced-ordered modeling in combination with logical rules was the foundation of much of the work presented in this thesis.

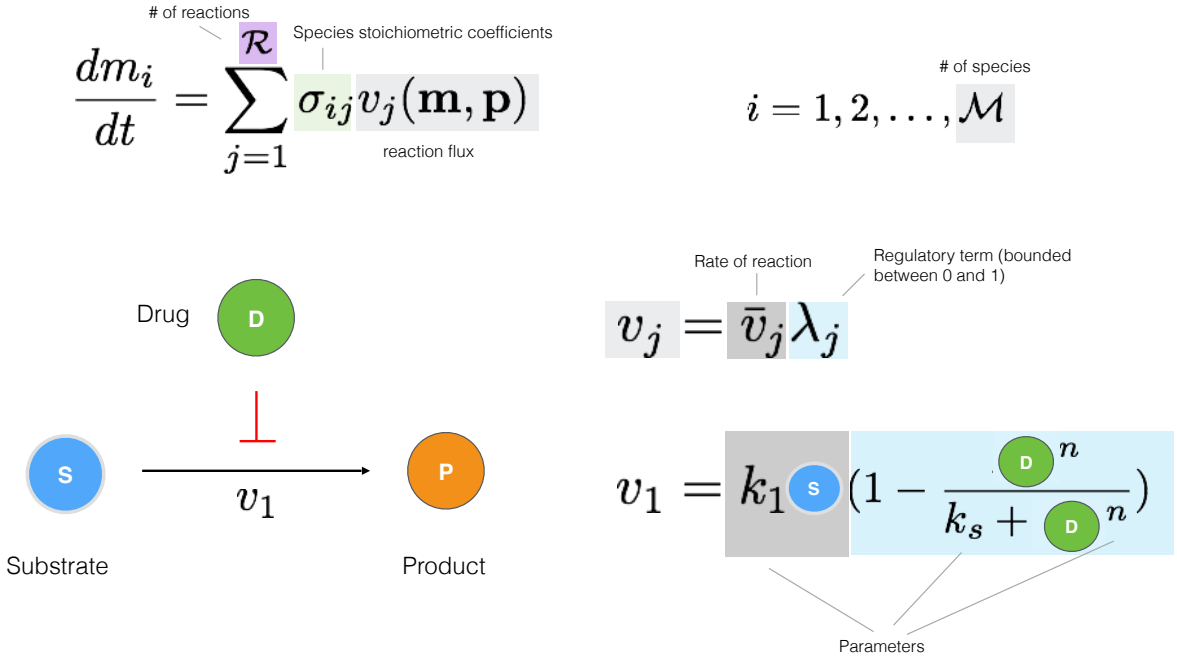

Figure 1.1: Schematic of how a system of ordinary differential equations (ODEs) are implemented to represent a network of biological interactions. The rate of change with respect to time of each individual species is represented by a mass balances. The rate of accumulation or depletion of the species are determined by the rate of reactions. The diagram shows an example of a substrate (S) being converted to product (P) at the rate of v_1, while the presence of drug (D) inhibits the reaction.

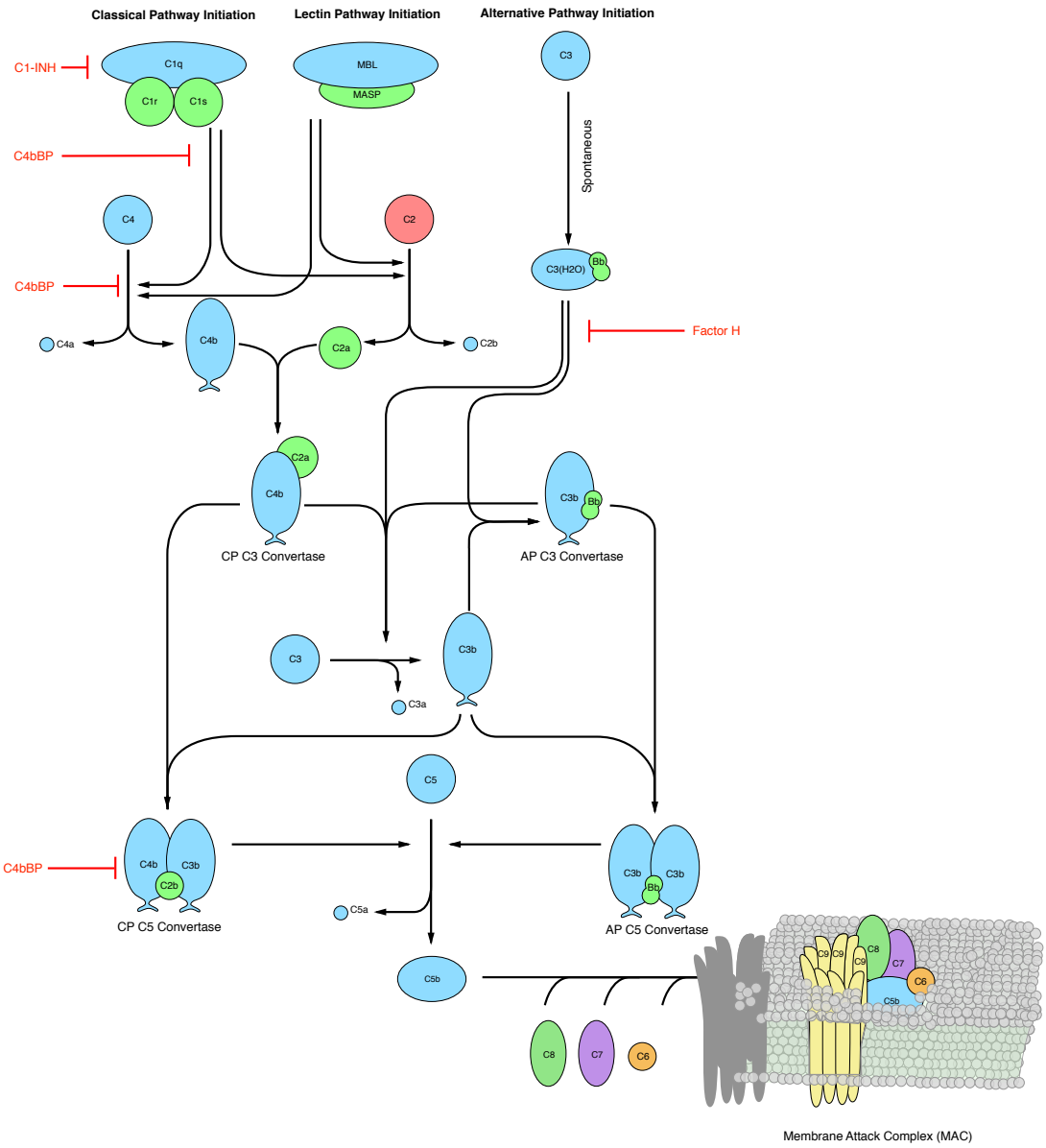

Figure 1.2: Simplified schematic of the human complement system. The complement cascade is activated through three pathways: the classical, the lectin, and the alternative pathways. Complement initiation results in the formation of classical or alternative C3 convertases, which amplify the initial complement response and signal to the adaptive immune system by cleaving C3 into C3a and C3b. C3 convertases further react to form C5 convertases which catalyze the cleavage of the C5 complement protein to C5a and C5b. C5b is critical to the formation of the membrane attack complex (MAC), while C5a recruits an adaptive immune response.

1.1.2 Constraint-based Modeling

As discussed above, there are many existing mathematical approaches to characterize biochemical networks such as cellular metabolism and its regulation [26, 71, 51, 8, 152, 180, 138, 186, 36, 247]. However, many of these methods require detailed kinetic and concentration information that are difficult or even impossible to obtain. Even with the growing data bases for cellular components and development of new reduced order modeling approaches [198], the applications of many mechanistic modeling methods are limited by the lack of kinetic and concentration information.

To overcome the lack of kinetic information, alternative course-grained modeling approaches have been used to study flux distributions of metabolic networks. In the post-genomics era, large-scale stoichiometric reconstructions of microbial metabolism popularized by static, constraint-based modeling techniques such as flux balance analysis (FBA) have become standard tools [146]. Since the first genome-scale reconstruction of *Escherichia coli* MG1655 by Edwards and Palsson [62], the reconstruction of over 100 organisms has followed [69]. These organisms included prokaryotes such as *E. coli* [68] or *B. subtilis* [182] were highly sought after in bioprocessing to maximize yields of desired products. More recently, network reconstructions have been completed on humans as well [55, 226]. FBA is a course-grained modeling approach that relies on a pseudo-steady-state assumption to reduce unidentifiable kinetic models into an underdetermined linear algebraic system that can be solved efficiently even for large systems. Traditionally, FBA models lack descriptions of metabolic regulation and control mechanisms that cells need for adapting to different chemical and environmental stimulus. FBA models instead chooses pathways in the net-

work by prescribing an objective function on metabolism. The use of an objective function is crucial for FBA models since predictions are highly dependent on the objective function used for the analysis. Common objective functions include maximizing growth [146, 161, 24, 61, 230], ATP production [191], and the production of desired products [233]. The utilization of objective function in FBA models is a huge advantage over mechanistic models because with little to no kinetic parameters, the maximization of biomass production allows for a wide range of predictions that are consistent with experimental observations for microbial systems [146, 161, 199, 59, 108, 41, 60, 24].

In the last decade, FBA has increased in accuracy and utility by incorporating additional biological knowledge from regulatory constraints, thermodynamics, and alternative classes of objective functions [130]. Second generation FBA models initially approached regulatory constraints as Boolean logic operators [41] that arise based on environmental cues. Palsson and coworkers took a bioinformatics approach and incorporated transcription and translation through the use of the E-Matrix to regulate the metabolism [225]. Meanwhile other second generation FBA models focused on incorporating constraints at a physiological level through the use of crowding constraints, enzyme solubility, or membrane economics that constrain the upper limit on the sum of certain or all the flux vectors [10, 210, 262]. The accuracy of the prediction from constraint-based models such as FBA is based heavily on the constraints implemented to reduced the solution space. Second generation FBA models introduce additional constraints that reshape and reduce the solution space to redirect the flux distribution guided by the objective function in a way that more accurately describe biological behaviors.

Recently, constraint-based modeling like flux balance analysis has been implemented to optimize the production of small molecules or biologics. Genome-scale reconstructions for human has been achieved by Palsson and coworkers [56] and adopted in studies of cancer metabolism [210, 263, 204]. In one of our previous work, Multi-scale computational study of the Warburg effect, reverse Warburg effect and glutamine addiction in solid tumors, we took a multi-scale modeling approach to describe the intracellular, cellular,and multicellular behaviors of cells within a tumor (Fig. 1.4). With this framework, we investigated the following hypotheses: Warburg Effect/Aerobic glycolysis, Reverse Warburg, and Glutamine Addiction. We begin by translating hypotheses from experimental studies into constraints and objectives within the FBA (Fig.1.4 A). We used FBA to obtain the yield coefficients (Y = maximum growth rate/flux of metabolite) for use in Monod-like kinetics of cellular growth at the individual cell level (Fig.1.4B). Finally, we simulated the growth dynamics of these cells at the multicellular scale to elucidate the implications of these metabolic scenarios (Fig.1.4BC). We addressed the impact of the metabolic phenotypes implied by current hypotheses on the growth dynamics of tumor cells in the resource-limited microenvironments that emerge after tumor initiation. This modeling framework opened a new avenue to explore the impact of different metabolic profiles at a tissue-scale.

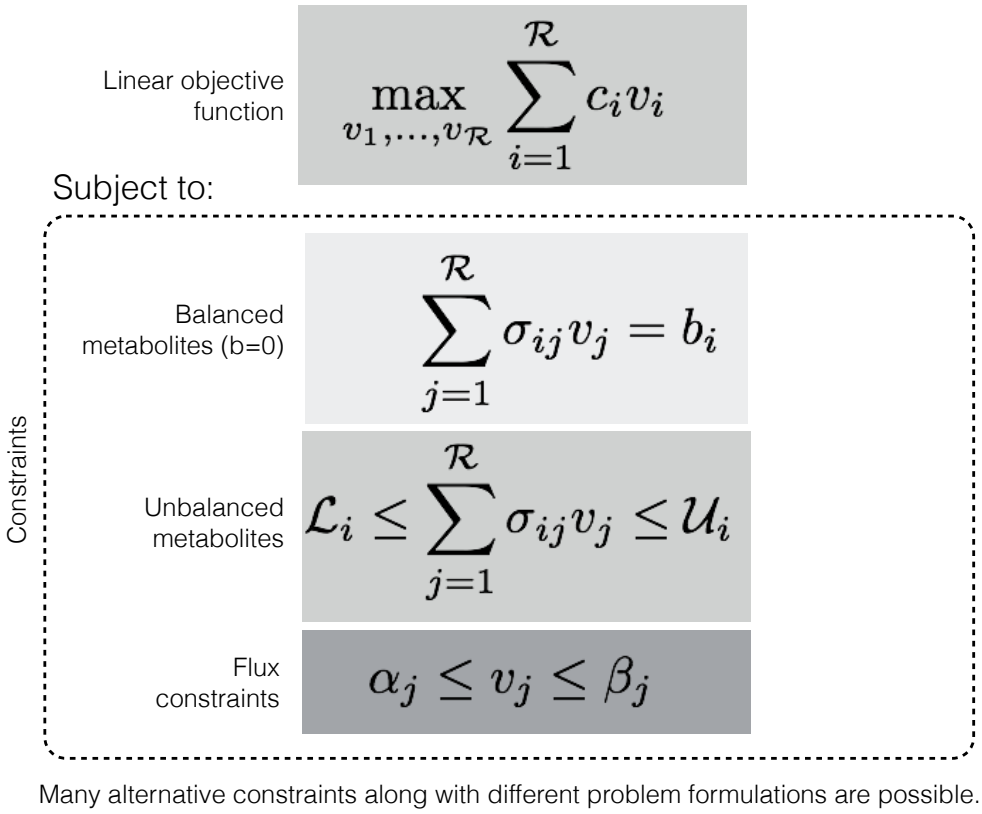

Figure 1.3: Flux balance analysis is formulated as a linear programming problem with an objective function to be maximized or minimized subjected to a set of constraints on the fluxes and the rate of accumulation of depletion.

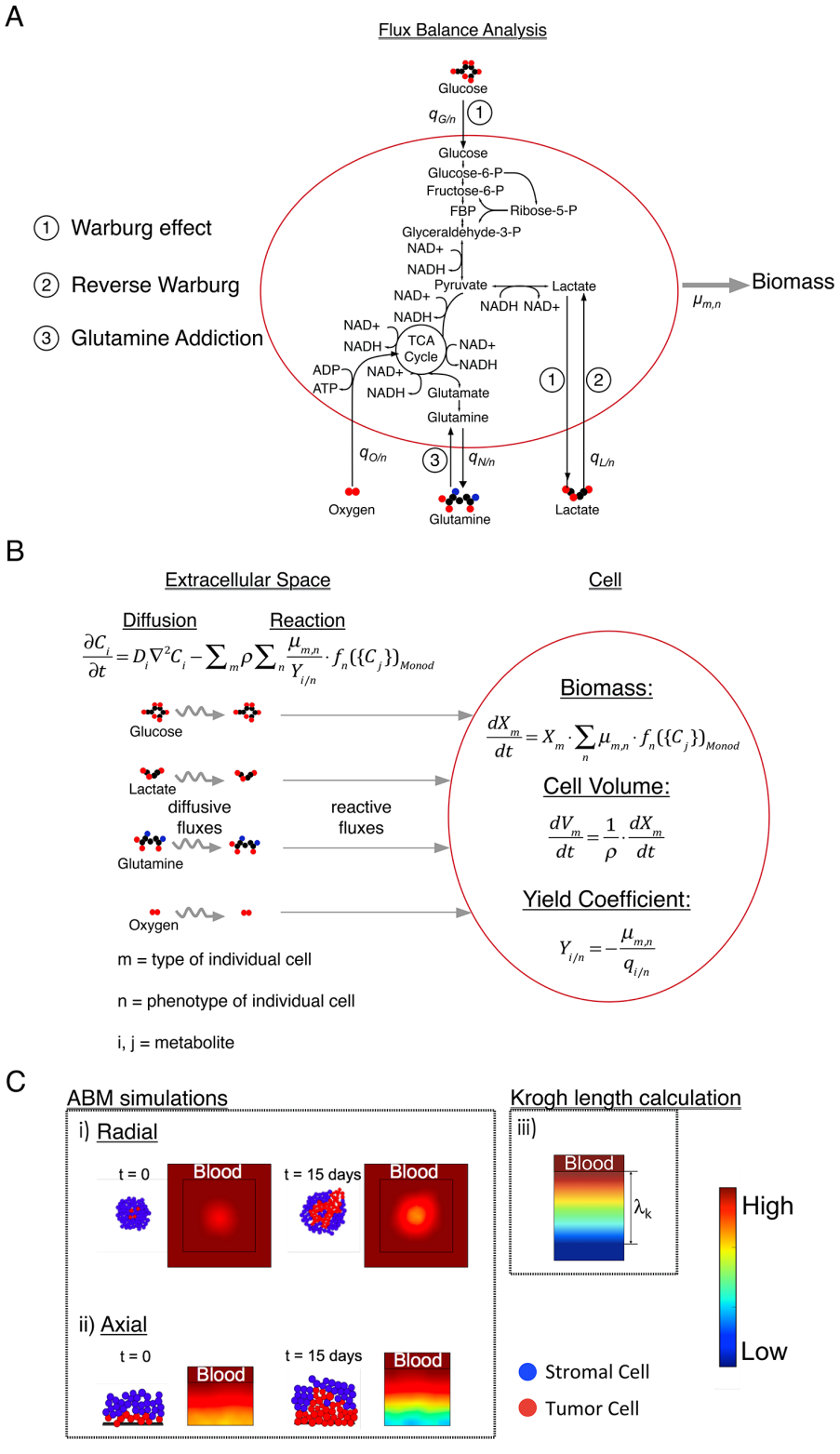

Figure 1.4: (Caption next page.)

Figure 1.4: (Previous page.) Multi-scale modeling of cancer metabolism [204]. (A) Flux Balance Analysis (FBA). The arrows represent fluxes of species within a reduced representation of cell metabolism and cell growth. Key steps associated with three hypotheses are labeled: (1) Warburg effect, (2) Reverse Warburg and (3) Glutamine addiction. The uptake and production rates are qi/n [g/g-DW-hr] for the ith metabolite and the nth metabolic phenotype. We impose the maximum growth rate, ,n [hr-1], for a given metabolic phenotype of each cell type as an objective function within the FBA. (B) Cell: Biomass (Xm [g]) growth of each cell type is modeled as a Monod-like process parameterized by the same maximum growth rates used in FBA that are modulated by functions of metabolite concentrations, fn(Cj)Monod. The change in Volume of the cell (Vm [L]) is calculated from biomass growth by applying a constant density of the cell ([g-DW/L]). Yield coefficients (Yi/n [g-DW/g]) for each metabolite (i) and corresponding metabolic phenotype (n) are defined in terms of the uptake and production rates (qi/n) obtained from FBA. Extracellular space: Species balances for each explicit metabolite follow reaction-diffusion kinetics and govern the concentration profiles of metabolite at the multicellular scale. These equations (Eqs 1-4 in text) are integrated into and solved within an agent-based model (ABM; iDynoMiCS). (C) ABM simulations: i) Radial, two dimensional growth: Tumor cells grow radially out from an initial cluster of cells with metabolites supplied at the edge of the cell mass such that radial gradients of concentration emerge (color map;red is high and blue is low concentration). Two phenotypes are displayed (red: tumor cells and blue: stromal cells). As the tumor grows, concentration gradients of metabolites become significant, making the tumor growth a diffusion-limited process that can result in different growth dynamics as well as distinct spatial distribution of cell subpopulations. ii) Axial, one dimensional growth: Layers of tumor cells (red) and stromal cells (blue) are initiated near a blood vessel that supplies metabolites (from the top), such as glucose and oxygen in the blood stream. Growth pushes cells deeper into the tissue, away from the vessel, such that strong gradients of metabolite can again occur. iii) Krogh length calculation: To evaluate the impact of diffusion limitations in a simple model, we treat cells as continuum with uniform, zeroth order kinetics of metabolite consumption to calculate the distance over which the concentration of limiting metabolites falls to zero within the tumor mass; we refer to this distance as the Krogh length of a given metabolite.

1.2 Current Approach in Cancer Therapy

Cancer is an umbrella term that encompasses a collection of related diseases. Douglas Hanahan and Robert Weinburg characterized that the hallmarks of cancer comprise six biological capabilities acquired during the multistep development of human tumors [87].They include sustaining proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, and activating invasion and metastasis.

The National Cancer Institute divides cancer into solid and liquid cancers. At the tissue level, solid cancers can aggregate and form solid tumours, which can be benign and malignant. Benign tumors can grow large, however they can be surgically removed and will not return, whereas malignant tumors sometimes do. As opposed to benign tumors, malignant tumors grow they can metastasize by breaking off and traveling to distant sites in the body through the blood or the lymph system and form new tumors far from the original tumor. At the tissue level, therapeutic interventions are aimed to prevent tumors from metastasizing with anti-angiogentic compounds that targets Vascular endothelial growth factor and Vascular endothelial growth factor receptors.

At the intracellular level, there are a multitude of combinations of genetic mutation that can contribute to the hallmarks of cancer. For instance tumor cells alter their metabolism to maintain unregulated cellular proliferation and survival, but this transformation leaves them reliant on constant supply of nutrients and energy [231]. In addition to dysregulated glucose metabolism for growth, accumulating evidences suggest that utilization of amino acids and lipids contributes significantly to cancer cell metabolism [231].

There currently exist a wide range of clinical interventions that targets cancers in different stages with different modalities. The most common treatments involve chemotherapy. Most of the chemotherapeutic agents available for cancer treatment function by interfering with cell division or DNA synthesis. Based on their mechanism of action (MOA), chemotherapeutic agents can be categorized as alkylating agents, antimetabolites, anthracyclines, topoisomerase inhibitors and plant alkaloids. Due to the detrimental side effects and subtypes of cancer that are not effect by tradition chemotherapies, there have been a great amount of effort devoted into developing targeted therapies.

Targeted therapies can modulate a variety of biological functions. Hormone therapies attempt to inhibit the growth of hormone-sensitive tumors such as prostate cancer [1]. Hormone therapies can work by inhibiting the native production of the hormone or interfering with the action of the hormone post production. However, cancers that are initially sensitive to hormone therapy can develop resistance and become unresponsive. For instance, Rogers and coworkers leveraged computational modeling to study castrate resistance prostate cancer [129]. Other modalities of therapy are required when hormone therapy is not responsive or not applicable to the tumor type. Cells uses signal transduction networks to relay information sensed from their surrounding, whether chemical or physical. However in cancer, often times the receptors that sense chemical or physical cues, or associated signaling molecules, become mutated and deregulated leading to a constant activation. Signal transduction inhibitors attenuates signaling molecules from activating downstream pathways that regulate the expression of oncogenes or tumor suppressors. Gene expression modulators modify the function of proteins that play a role in controlling gene expression. There are many other therapies that targets different aspects of disease progression

such as: angiogeneis inhibitors, immunotherapies, antibodies, apoptosis inducers, and cancer vaccines. Recently, there has been a growth in the development of targeted therapies aimed at the metabolism.

1.3 Cancer Metabolism

Cancer cells are known to rewire their metabolism to promote growth, survival, and long-term maintenance. Although cancer is a genetically diverse disease, the common feature of this altered metabolism is increased glucose uptake, and fermentation of glucose to lactate in the presence of oxygen persists. This phenomenon was first observed by Otto Warburg in the 1920s, who later proposed that dysfunctional mitochondria is the root of aerobic glycolysis [243] and further hypothesized that this event is the primary cause of cancer. This phenomenon, later known as the Warburg Effect, is thought to be required for tumor growth [65, 209]. Until recently, the majority of studies of cancer metabolism focused on glucose metabolism, specifically the increased conversion of glucose to lactate. In addition to glycolysis, this observation showed that mammalian cells rely on both glucose and glutamine, and some cancer cells also exhibited addiction to glutamine despite the fact that it is a nonessential amino acid that can be synthesized from glucose. The high rate of glutamine uptake exhibited by glutamine-dependent cells does not appear to result solely from its role as a nitrogen donor in nucleotide and amino acid biosynthesis. In addition to its role in replenishing carbon in the TCA cycle, glutamine also contributed to a wide range of pathways in cells, producing amino acids, nucleotides and fatty acids, as well as playing an important role in reactive oxygen species (ROS) homeostasis, mTOR activation, and the hexosamine biosynthesis pathway [216].

Metabolism is regulated by signaling events upstream that informs cellular decision making. For example, phosphoinositide 3-kinase (PI3K) signaling is linked to both growth control and glucose metabolism. The PI3K/AKT/mTOR pathway is an intracellular signal transduction pathway that promotes metabolism, proliferation, cell survival, growth and angiogenesis in response to extracellular signals. This is mediated through serine and/or threonine phosphorylation of a range of downstream substrates. Key proteins involved are phosphatidylinositol 3-kinase (PI3K) and AKT (Protein Kinase B). The PI3K/AKT/mTOR pathway has many downstream effects and is carefully regulated. One important regulatory point of the PI3K-AKT pathway is the PI3K phosphorylation of phosphatidylinositol (4,5)-bisphosphate (PIP2) to phosphatidylinositol (3,4,5)-trisphosphate (PIP3) which in turn activates AKT. Phosphatase and tensin homolog (PTEN) is a mainregulator which converts PIP3 into PIP2. Simiarily, protein phosphatase 2A (PP2A), which dephosphorylates AKT at Thr308, and phosphatase PHLPP, which dephosphorylates Akt at Ser473, are also two negative regulation proteins. Upon activation (phosphorylated at Thr308 and Ser473), AKT can be translocated from the plasma membrane to activate its downstream targets such as mTOR that can effect cell cycle progression, proliferation, apoptosis, and glucose metabolism. Recent studies have shown other interactions between signaling and metabolism go beyond the traditional understanding (ie. signaling induces gene expression that alters the metabolism). For instance, AKT has recently been shown to directly activate hexokinase II by phosphorylation at Thr473, leading to increased mitochondria bound hexokinase II. Interestingly, AKT also interacts with phosphofructose kinase 1 platelet isoform by phosphorylating PFKP at Ser386. This phosphorylation inhibits the binding of TRIM21 E3 ligase to PFKP and the subsequent

ubiquitylation and degradation of PFKP [194, 17].

1.4 HL60 differentiation

Differentiation induction chemotherapy (DIC), using agents such as the vitamin A derivative all-trans retinoic acid (ATRA), is a promising approach for the treatment of many cancers [25, 222, 30, 229, 148]. For example, ATRA treatment induces remission in 80–90% of promyelocytic leukemia (APL) PML-RARα-positive patients [178], thereby transforming a fatal diagnosis into a manageable disease [37]. However, remission is sometimes not durable and relapsed cases exhibit emergent ATRA resistance [244, 73]. To understand the basis of this resistance, we must first understand the ATRA-induced differentiation program. Toward this challenge, lessons learned in model systems, such as the lineage-uncommitted human myeloblastic cell line HL-60 reported to closely resemble patient derived cells [90], could inform our analysis of the differentiation programs occurring in patients. Patient derived HL-60 leukemia cells have been a durable experimental model since the 1970's to study differentiation [18]. HL-60 undergoes cell cycle arrest and either myeloid or monocytic differentiation following stimulation; ATRA induces G1/G0-arrest and myeloid differentiation in HL-60 cells, while 1,25-dihydroxy vitamin D3 (D3) induces arrest and monocytic differentiation. Commitment to cell cycle arrest and differentiation requires approximately 48 hr of treatment, during which HL-60 cells undergo two division cycles.

Sustained mitogen-activated protein kinase (MAPK) pathway activation is a defining feature of ATRA-induced HL-60 differentiation. ATRA drives sus-

tained activation of the Raf/MEK/ERK pathway, leading to arrest and differentiation [254]. Betraying a feedback loop, MEK inhibition results in the loss of ERK as well as Raf phosphorylation and the failure to arrest and differentiate in response to ATRA [101]. Retinoic acid (and its metabolites) are ligands for the hormone activated nuclear transcription factors retinoic acid receptor (RAR) and retinoid X receptor (RXR) [164]. RAR/RXR activation is necessary for ATRA-induced Raf phosphorylation [101] in concert with the formation of an ATRA-induced signalsome complex at the membrane, which drives MAPK activation. While the makeup of the signalsome complex is not yet known, we do know that it is composed of Src family kinases Fgr and Lyn, PI3K, c-Cbl, Slp76, and KSR, plus transcription factors AhR and IRF1 [33, 205, 208, 255, 165]. Signalsome activity is driven by ATRA-induced expression of CD38 and putatively the heterotrimeric Gq protein-coupled receptor BLR1 [32, 241]. Members of the BLR1 transcriptional activator complex, e.g. NFATc3 and CREB, are phosphorylated by ERK, JNK or p38 MAPK family members suggesting positive feedback between the signalsome and MAPK activation [251, 150]. BLR1 overexpression enhanced Raf phosphorylation and accelerated terminal differentiation, while Raf inhibition reduced BLR1 expression and ATRA-induced terminal differentiation [242]. The expression of signalsome components, e.g. BLR1, was Raf dependent, while Raf activation depended upon the signalsome. A previous computational study of ATRA-induced differentiation of HL-60 cells suggested that the BLR1-MAPK positive feedback circuit was sufficient to explain ATRA-induced sustained MAPK activation, and the expression of a limited number of functional differentiation markers [223]. Model analysis also suggested that Raf was the most distinct of the MAPK proteins. However, this previous study developed and analyzed a complex model, thus leaving open the critical ques-

tion of what is the minimal positive feedback circuit required to drive ATRA-induced differentiation. The activation of ATRA bound RARα/RXR regulates the expression of the transcription factors CCATT/enhancer binding protein α (C/EBPα), PU.1, and Egr-1. The transcription factors PU.1 and C/EBPα are master regulators and have positive effects on both granuolytic (ATRA) and monocytic (vitamin D3) maturation, however their ratio determines granulocytic versus monocyctic lineage selection. Similarly, two antagonistic repressors, Gfi-1 and EGR-1 lie directly downstream of PU.1 and C/EBPα. Gfi-1 represses monocytic differentiation and promotes granulocytic different ion, while EGR-1 acts conversely. In turn, these transcription factors, in combination with cRaf-pS621, regulates the expression of downstream phenotypic markers such as CD38, CD11b or p47Phox.

1.5 Outlook

Although cancer is a genetically diverse disease, a common feature between the different type of cancers is altered metabolism. Cancer cells rewire their metabolism to uptake more glucose to produce lactate as an inefficient way of energy production. We believe that metabolism can be used to better understand the deregulation of signaling and gene expression. There is a great opportunity to utilize a systems approach to evaluate combination therapy targeting consisting of different modalities and modes of action. In this work, we leveraged our current understanding of signaling transduction, and metabolic networks to develop a systems approach to quantitatively unravel the mechanisms that regulate the hallmarks of cancer. More specifically, we combined a top-down approach starting from the signaling, and a bottom-up approach from

the metabolism to better understand how the cells respond to environmental cues.

In Chapter 2, we utilized our understanding of ATRA-induced differentiation programming in HL-60 cells to develop computational tools that can integrate signal transduction networks with transcription factor networks. In Chapter 3, we expanded on this work and explored combination treatment of ATRA and D3 to further investigate the transcription factor network and treatment specific pathway sensitivity. In Chapter 4, we describe the experimental exploration of the combination therapies using the SRC family kinase, Bosutinib and ATRA in HL-60 cells. In Chapter 5, we transitioned our focus towards metabolism. We developed a novel constraint-based model that eliminated the pseudo-steady state assumption and captured metabolite dynamics. In addition, we used this modeling framework to determine the metabolite measurements that contributed to the best model performance. In Chapter 6, we leveraged our understanding of metabolite selection from our previous studies and developed analytical techniques to quantify metabolite abundance. Finally, chapter 7 concludes with future work.

CHAPTER 2

AN EFFECTIVE MODEL OF THE RETINOIC ACID INDUCED HL-60 DIFFERENTIATION PROGRAM

2.1 Abstract

[1] In this study, we present an effective model All-Trans Retinoic Acid (ATRA)-induced differentiation of HL-60 cells. The model describes reinforcing feedback between an ATRA-inducible signalsome complex involving many proteins including Vav1, a guanine nucleotide exchange factor, and the activation of the mitogen activated protein kinase (MAPK) cascade. We decomposed the effective model into three modules; a signal initiation module that sensed and transformed an ATRA signal into program activation signals; a signal integration module that controlled the expression of upstream transcription factors; and a phenotype module which encoded the expression of functional differentiation markers from the ATRA-inducible transcription factors. We identified an ensemble of effective model parameters using measurements taken from ATRA-induced HL-60 cells. Using these parameters, model analysis predicted that MAPK activation was bistable as a function of ATRA exposure. Conformational experiments supported ATRA-induced bistability. Additionally, the model captured intermediate and phenotypic gene expression data. Knockout analysis suggested Gfi-1 and PPARγ were critical to the ATRA-induced differentiation program. These findings, combined with other literature evidence, suggested that reinforcing feedback is central to hyperactive signaling in a diversity of cell fate programs.

[1]This work is published and adopted from Scientific Reports https://doi.org/10.1038/s41598-017-14523-5

2.2 Introduction

Differentiation induction chemotherapy (DIC), using agents such as the vitamin A derivative all-trans retinoic acid (ATRA), is a promising approach for the treatment of many cancers [25, 222, 30, 229, 148]. For example, ATRA treatment induces remission in 80–90% of promyelocytic leukemia (APL) PML-RARα-positive patients [178], thereby transforming a fatal diagnosis into a manageable disease [37]. However, remission is sometimes not durable and relapsed cases exhibit emergent ATRA resistance [244, 73]. To understand the basis of this resistance, we must first understand the ATRA-induced differentiation program. Toward this challenge, lessons learned in model systems, such as the lineage-uncommitted human myeloblastic cell line HL-60 reported to closely resemble patient derived cells [90], could inform our analysis of the differentiation programs occurring in patients. Patient derived HL-60 leukemia cells have been a durable experimental model since the 1970's to study differentiation [18]. HL-60 undergoes cell cycle arrest and either myeloid or monocytic differentiation following stimulation; ATRA induces G1/G0-arrest and myeloid differentiation in HL-60 cells, while 1,25-dihydroxy vitamin D3 (D3) induces arrest and monocytic differentiation. Commitment to cell cycle arrest and differentiation requires approximately 48 hr of treatment, during which HL-60 cells undergo two division cycles.

Sustained mitogen-activated protein kinase (MAPK) pathway activation is a defining feature of ATRA-induced HL-60 differentiation. ATRA drives sustained activation of the Raf/MEK/ERK pathway, leading to arrest and differentiation [254]. Betraying a feedback loop, MEK inhibition results in the loss of ERK as well as Raf phosphorylation and the failure to arrest and differenti-

ate in response to ATRA [101]. Retinoic acid (and its metabolites) are ligands for the hormone activated nuclear transcription factors retinoic acid receptor (RAR) and retinoid X receptor (RXR) [164]. RAR/RXR activation is necessary for ATRA-induced Raf phosphorylation [101] in concert with the formation of an ATRA-induced signalsome complex at the membrane, which drives MAPK activation. While the makeup of the signalsome complex is not yet known, we do know that it is composed of Src family kinases Fgr and Lyn, PI3K, c-Cbl, Slp76, and KSR, plus transcription factors AhR and IRF1 [33, 205, 208, 255, 165]. Signalsome activity is driven by ATRA-induced expression of CD38 and putatively the heterotrimeric Gq protein-coupled receptor BLR1 [32, 241]. BLR1 (also known as CXCR5), identified as an early ATRA (or D3)-inducible gene using differential display [252], is necessary for MAPK activation and differentiation [241], and drives signalsome activity. Studies of the BLR1 promoter identified a non-canonical RARE site consisting of a 17 bp GT box approximately 1 kb upstream of the transcriptional start that conferred ATRA responsiveness [241]. Members of the BLR1 transcriptional activator complex, e.g. NFATc3 and CREB, are phosphorylated by ERK, JNK or p38 MAPK family members suggesting positive feedback between the signalsome and MAPK activation [251, 150]. BLR1 overexpression enhanced Raf phosphorylation and accelerated terminal differentiation, while Raf inhibition reduced BLR1 expression and ATRA-induced terminal differentiation [242]. In particular, Raf phosphorylation of the NFATc3 transcription factors at the BLR1 promoter enables transcriptional activation at the RARE by ATRA bound to RAR/RXR [80]. BLR1 knock-out cells failed to activate Raf or differentiate in the presence of ATRA [242]. Interestingly, both the knockdown or inhibition of Raf, also reduced BLR1 expression and functional differentiation [242]. Thus, the expression of signalsome components e.g., BLR1

was Raf dependent, while Raf activation depended upon the signalsome. A previous computational study of ATRA-induced differentiation of HL-60 cells suggested that the BLR1-MAPK positive feedback circuit was sufficient to explain ATRA-induced sustained MAPK activation, and the expression of a limited number of functional differentiation markers [223]. Model analysis also suggested that Raf was the most distinct of the MAPK proteins. However, this previous study developed and analyzed a complex model, thus leaving open the critical question of what is the minimal positive feedback circuit required to drive ATRA-induced differentiation.

In this study, we explored this question using a minimal mathematical model of the key architectural feature of ATRA induced differentiation of HL-60 cells, namely positive feedback between an ATRA-inducible signalsome complex and MAPK activation. The ATRA responsive signalsome-MAPK circuit was then used to drive a downstream gene expression program which encoded for the expression of intermediate and functional differentiation markers. The effective model used a novel framework which integrated logical rules with kinetic modeling to describe gene expression and protein regulation, while largely relying upon biophysical parameters from the literature. This formulation significantly reduced the size and complexity of the model compared to the previous study of Tasseff et al., while increasing the breadth of the biology described [223]. The effective model, despite its simplicity, captured key features of ATRA induced differentiation of HL-60 cells. Model analysis predicted the bistability of MAPK activation as a function of ATRA exposure; conformational experiments supported ATRA-induced bistability. Model simulations were also consistent with measurements of the influence of MAPK inhibitors, and the failure of BLR1 knockout cells to differentiate when exposed to ATRA. In addition, the expres-

sion of intermediate and phenotypic differentiation markers as also captured following ATRA exposure. Lastly, we showed through immunoprecipitation and inhibitor studies, that the guanine nucleotide exchange factor Vav1 is potentially a new ATRA-inducible member of the signalsome complex functioning as a regulator that contributes to signal amplification in the signalsome. Taken together, these findings when combined with other literature evidence, suggested that reinforcing feedback was central to differentiation programs generally, and necessary for ATRA-induced differentiation. The model answers a biologically important question that is not easily experimentally attacked, namely given the complexity of the signaling machine and the pathways it embodies, is there a critical small suite of molecules that are the action elements seminal to eliciting ATRA-induced cell differentiation and G0 arrest.

2.3 Results

We constructed an effective model of ATRA-induced HL-60 differentiation which described signaling and gene expression events following the addition of ATRA (Fig. 5.1). HL-60 is a NCI-60 cell model that is widely used model for studying the mechanism of action of ATRA. There is an immense body of literature on HL-60, therefore it was chosen for modeling. For molecules where there was not explicit HL-60 literature, then literature for well-known interactions from other experimental systems was curated to fill lacuna for the modeling. The model connectivity was developed from literature and the studies presented here (Table 3.1). We decomposed the ATRA program into three modules; a signal initiation module that sensed and transformed the ATRA signal into activated cRaf-pS621 and the ATRA-RAR/RXR (Trigger) signals (Fig. 5.1A); a

signal integration module that controlled the expression of upstream transcription factors given cRaf-pS621 and activated Trigger signals (Fig. 5.1B); and a phenotype module which encoded the expression of functional differentiation markers from the ATRA-inducible transcription factors (Fig. 5.1C). In particular, Trigger (a surrogate for the RARα/RXR transcriptional complex) regulated the expression of the transcription factors CCATT/enhancer binding protein α (C/EBPα), PU.1, and Egr-1. In turn, these transcription factors, in combination with cRaf-pS621, regulated the expression of downstream phenotypic markers such as CD38, CD11b or p47Phox. Each component of these modules was described by a mRNA and protein balance equation. Additionally, the signal initiation module also described the abundance of activated species e.g., Trigger and cRaf-pS621 whose values were derived from unactivated Trigger and cRaf protein levels. Lastly, because the population of HL-60 cells was dividing, we also considered a dilution term in all balance equations. The signal initiation module contained nine differential equations, while the signal integration and phenotype modules were collectively encoded by 54 differential equations. Model parameters were taken from literature (Table 2.2), or estimated from experimental data using heuristic optimization (see materials and methods).

The signal initiation module recapitulated sustained signalsome and MAPK activation following exposure to 1μM ATRA (Fig. 2.2A-B). An ensemble of effective model parameters was estimated by minimizing the difference between simulations and time-series measurements of BLR1 mRNA and cRaf-pS621 following the addition of 1μM ATRA. We focused on the S621 phosphorylation site of cRaf since enhanced phosphorylation at this site is a defining characteristic of sustained MAPK signaling activation in HL-60. The effective model captured both ATRA-induced BLR1 expression (Fig. 2.2A) and sustained phosphoryla-

tion of cRaf-pS621 (Fig. 2.2B) in a growing population of HL-60 cells. Together, the reinforcing feedback within the signalsome and its embedded MAPK signaling axis led to sustained activation over multiple cellular generations. However, the effective model failed to capture the decline of BLR1 message after 48 hr of ATRA exposure. This suggested that we captured the logic leading to the onset of differentiation, but failed to describe program shutdown. Much of the focus in the literature has been on understanding the initiation of differentiation, with little attention paid to understanding how a program is terminated. This is a potential new direction that could be explored. Next, we tested the response of the signal initiation module to different ATRA dosages.

The signal initiation model was bistable with respect to ATRA induction (Fig. 2.2C-D). Phaseplane analysis predicted two stable steady-states when ATRA was present below a critical threshold (Fig. 2.2C), and only a single steady-state above the threshold (Fig. 2.2D). In the lower stable state, neither the signalsome nor cRaf-pS621 were present (thus, the differentiation program was inactive). However, at the higher stable state, both the signalsome and cRaf-pS621 were present, allowing for sustained activation and differentiation. Interestingly, when ATRA was above a critical threshold, only the activated state was accessible (Fig. 2.2D). To test these findings, we first identified the ATRA threshold. We exposed HL-60 cells to different ATRA concentrations for 72 hr (Fig. 2.2E). Morphological changes associated with differentiation were visible for ATRA $\geq 0.25\ \mu$M, suggesting the critical ATRA threshold was near this concentration. Next, we conducted ATRA washout experiments to determine if activated cells remained activated in the absence of ATRA. HL-60 cells locked into an activated state remained activated following ATRA withdraw (Fig. 2.3C). This sustained activation resulted from reinforcing feedback between the sig-

nalsome and the MAPK pathway. Thus, following activation, if we inhibited or removed elements from the signal initiation module we expected the signalsome and MAPK signals to decay. We simulated ATRA induced activation in the presence of kinase inhibitors, and without key circuit elements. Consistent with experimental results using multiple MAPK inhibitors, ATRA activation in the presence of MAPK inhibitors lowered the steady-state value of signalsome (Fig. 2.3A). In the presence of BLR1, the signalsome and cRaf-pS621 signals were maintained following ATRA withdraw (Fig. 2.3B, gray). On the other hand, BLR1 deletion removed the ability of the circuit to maintain a sustained MAPK response following the withdraw of ATRA (Fig. 2.3B, blue). Lastly, washout experiments in which cells were exposed to 1μM ATRA for 24 hr, and then transferred to fresh media without ATRA, confirmed the persistence of the self sustaining activated state for up to 144 hr (Fig. 2.3C). Thus, these experiments confirmed that reinforcing positive feedback likely drives the ATRA-induced differentiation program. Next, we analyzed the ATRA-induced downstream gene expression program following signalsome and cRaf activation.

The signal integration and phenotype modules described ATRA-induced gene expression in wild-type HL-60 cells (Fig. 2.4). The signal initiation module produced two outputs, activated Trigger and cRaf-pS621 which drove the expression of ATRA-induced transcription factors, which then in turn activated the phenotypic program. We assembled the connectivity of the signal integration and phenotypic programs driven by Trigger and cRaf-pS621 from literature (Table 3.1). We estimated the parameters for the signal initiation, and phenotype modules from steady-state and dynamic measurements of transcription factor and phenotypic marker expression following the addition of ATRA [117, 114, 116, 119]. However, the bulk of the model parameters were taken from

literature [172] and were not estimated in this study (see materials and methods). The model simulations captured the time dependent expression of CD38 and CD11b following the addition ATRA (Fig. 2.4A), and the steady-state for signal integration and phenotypic markers (Fig. 2.4B). Lastly, we used the *predicted* values of the p21 and E2F protein abundance to estimate a blackbox model of ATRA-induced G0 arrest (Fig. 2.5). The phenotype module predicted p21 expression significantly increased and E2F expression decreased, in response to ATRA exposure (Fig. 2.5A). We then used the ratio of these values in a polynomial model to calculate the fraction of HL-60 cells in G0 arrest following the addition of ATRA (Fig. 2.5B). The third-order polynomial model captured the trend in measured G0-arrest values as a function of time, and was robust to uncertainty in the measured data (Fig. 2.5B, gray). Taken together, the output of the signal integration and phenotypic modules was consistent with time-series and steady-state measurements, thereby validating the assumed molecular connectivity. Moreover, outputs from the phenotype module described the trend in ATRA-induced G0 cell cycle arrest. Next, we explored which proteins and protein interactions in the signal integration module most influenced the system response.

The Gfi-1 and PPARγ proteins were important regulators of ATRA-induced signal integration and phenotypic change (Fig. 2.6). We conducted pairwise gene knockout simulations in the signal integration and phenotype modules to estimate which proteins controlled the processing of the Trigger and cRaf-S621 signals. The difference between the system state with and without the gene knockouts (encoded as a normalized state displacement matrix) was decomposed using Singular Value Decomposition (SVD). A panel of ten parameter sets was sampled, and the average normalized displacement matrix was de-

composed. The first six modes (approximately 36% of the total) described >95% of the gene knockout variance, with the most important components of these modes being the Gfi-1 and PPARγ proteins, and to a lesser extent PU.1, C/EBPα and and AP1 (Fig. 2.6A). To better understand which protein-DNA connections were important, we simulated the pairwise deletion of interactions between these proteins and their respective regulatory targets. Singular value decomposition of the normalized state displacement matrix assembled from the pairwise connection deletions, suggested the first six modes (approximately 26% of the total) accounted for >90% of the variance. Globally, the most sensitive interactions controlled p47Phox and p21 expression, markers for cell-cycle arrest and reactive oxygen formation phenotypic axes activated following ATRA addition (Fig. 2.6B). While the p21 spot appeared small, it was the second highest ranked response behind p47Phox, in the largest response mode. The interactions associated with these shifts likely involved important components; the deleted interactions involved the action of PU.1, C/EBPα and cRaf at both the p47Phox and p21 promoters, as well as PPARγ action for p21. Taken together, the gene and interaction knockout studies showed that the action of PPARγ, Gfi-1 and C/EBPα was consistently important over multiple target genes. The connection knockout analysis also revealed robustness within the network. For example, no pair of deletions qualitatively changed the expression of regulators such as PU.1, Oct1, Oct4 or PPARγ. Thus, the expression of these species was robust to disturbance in the connectivity. To better understand the combined influence of the PPARγ and Gfi-1 deletions, we computed the fold change in the protein levels in the single (Gfi-1$^{-/-}$ or PPAR$\gamma^{-/-}$) and double (Gfi-1$^{-/-}$ and PPAR$\gamma^{-/-}$) mutants for the best fit parameter set (Fig. 2.7). Deletion of Gfi-1 led to a 2-4 fold increase in EGR-1, CD11b and C/EBPα expression, and a >8 fold increase

in PU.1 abundance (Fig. 2.7,blue). On the other hand, deletion of PPARγ led to >8 fold downregulation of CD38, p21, IRF1 and Oct1 (Fig. 2.7,red). Both knockouts slightly increased E2F expression, but neither influenced the expression of p47Phox. The double mutant was qualitatively similar to the combined behavior of the two single mutant cases. Taken together, Gfi-1 and PPARγ controlled the cell-cycle arrest and receptor signaling axes, with PPARγ regulating CD38, IRF1 and p21 expression while Gfi-1 controlled CD11b expression. These simulations suggested deletion of PPARγ and Gfi-1 would not interfere with reactive oxygen formation, but would limit the ability of HL-60 cells to arrest. However, this analysis did not give insight into which components upstream of the signal initiation module were important. Toward this question, we explored the composition and regulation of the signalsome complex by experimentally interrogating a panel of possible Raf interaction partners.

The full composition of the signalsome, and the kinase therein ultimately responsible for mediating ATRA-induced Raf activation is still not known. To explore this question, we conducted immunoprecipitation and subsequent Western blotting to identify interactions between Raf and 19 putative interaction partners. A panel of 19 possible Raf interaction partners (kinases, GTPases, scaffolding proteins etc) was constructed based upon known signaling pathways. We did not consider the most likely binding partner, the small GTPase RAS, as previous studies have ruled it out in MAPK activation in HL-60 cells [242, 126]. Total Raf was used as a bait protein for the immunoprecipitation studies. Interrogation of the Raf interactome thus suggested Vav1 was involved with ATRA-induced initiation of MAPK activity (Fig. 2.8). Western blot analysis using total Raf and Raf-pS621 specific antibodies confirmed the presence of the bait protein, total and phosphorylated forms, in the immunoprecipitate

(Fig. 2.8A). Of the 19 proteins sampled, Vav1, Src, CK2, Akt, and 14-3-3 co-precipitated with Raf, suggesting their co-existence in a complex was possible. However, only the associations between Raf and Vav1, and Raf and Src were ATRA-inducible (Fig. 2.8). The interaction between Vav1 and Raf was one of the most prominent interactions in the panel, and it was crippled by inhibiting Raf. Furthermore, the Vav1 and Src associations were correlated with Raf-pS621 abundance in the precipitate. Other proteins e.g., CK2, Akt and 14-3-3, generally bound Raf regardless of phosphorylation status or ATRA treatment. The remaining proteins sampled were expressed in whole cell lysate (Fig. 2.8B), but were not detectable in the immuno-precipitate with Raf IP; consistent with the potential importance of the Raf-Vav interaction for signaling, it paralleled Raf phosphorylation at S621, a putative telltale of the activated kinase. Furthermore, treatment with the Raf kinase inhibitor GW5074 following ATRA exposure reduced the association of both Vav1 with Raf and Src with Raf (Fig. 2.8C), although the signal intensity for Src was notably weak. However, GW5074 did not influence the association of CK2 or 14-3-3 with Raf, further demonstrating their independence from Raf phosphorylation. Interestingly, the Raf-Akt interaction qualitatively increased following treatment with GW5074; however, it remained unaffected by treatment with ATRA. Src family kinases are known to be important in myeloid differentiation [173] and their role in HL-60 differentiation has been investigated elsewhere [33]. Given the existing work and variable reproducibility in the context of the Raf immunoprecipitate, we did not investigate the role of Src further in this study. Taken together, the immunoprecipitation and GW5074 results implicated Vav1 association to be correlated with Raf activation following ATRA-treatment. Further, while we observed possible immunoprecipitation of Src with Raf, the western blot results showed inconsis-

tent results and significant non-specific binding; therefore we could not rule in or out a Src/Raf interaction. Previous studies demonstrated that a Vav1-Slp76-Cbl-CD38 complex plays an important role in ATRA-induced MAPK activation and differentiation of HL-60 cells [208]. Here we did not observe direct interaction of Raf with Cbl or Slp76; however, this interaction could could be involved upstream. Next, we considered the effect of the Raf kinase inhibitor GW5074 on functional markers of ATRA-induced growth arrest and differentiation.

Inhibition of Raf kinase activity modulated MAPK activation and differentiation markers following ATRA exposure (Fig. 2.8D-F). ATRA treatment alone statistically significantly increased the G1/G0 percentage over the untreated control, while GW5074 alone had a negligible effect on the cell cycle distribution (Fig. 2.8D). Surprisingly, the combination of GW5074 and ATRA statistically significantly increased the G1/G0 population (82 ± 1%) compared with ATRA alone (61 ± 0.5%). Increased G1/G0 arrest following the combined treatment with GW5074 and ATRA was unexpected, as the combination of ATRA and the MEK inhibitor (PD98059) has been shown previously to decrease ATRA-induced growth arrest [254]. However, growth arrest is not the sole indication of functional differentiation. Expression of the cell surface marker CD11b has also been shown to coincide with HL-60 cells myeloid differentiation [94]. We measured CD11b expression, for the various treatment groups, using immunofluorescence flow cytometry 48 hr post-treatment. As with G1/G0 arrest, ATRA alone increased CD11b expression over the untreated control, while GW5074 further enhanced ATRA-induced CD11b expression (Fig. 2.8E). GW5074 alone had no statistically significant effect on CD11b expression, compared with the untreated control. Lastly, the inducible reactive oxygen species (ROS) response was used as a functional marker of differentiated neutrophils [32]. We measured

the ROS response induced by the phorbol ester 12-O-tetradecanoylphorbol-13-acetate (TPA) using flow cytometry. Untreated cells showed no discernible TPA response, with only 7.0 ± 3.0% ROS induction (Fig. 2.8F). Cells treated with ATRA had a significantly increased TPA response, 53 ± 7% ROS induction 48 hr post-treatment. Treatment with both ATRA and GW5074 statistically significantly reduced ROS induction (22 ± 0.6%) compared to ATRA alone. Interestingly, Western blot analysis did not detect a GW5074 effect on ATRA-induced expression of p47Phox, a required upstream component of the ROS response (Fig. 2.8F, bottom). Thus, the inhibitory effect of GW5074 on inducible ROS might occur downstream of p47Phox expression. However, the ROS producing complex is MAPK dependent, therefore it is also possible that GW5074 inhibited ROS production by interfering with MAPK activation (in which case the p47Phox marker might not accurately reflect phenotypic conversion and differentiation).

2.4 Discussion

In this study, we presented an effective model of ATRA-inducible differentiation of HL-60 cells. The model consisted of three modules: a signal initiation module that sensed and transformed the ATRA signal into activated cRaf-pS621 and the ATRA-RAR/RXR (Trigger) signals; a signal integration module that controlled the expression of upstream transcription factors given cRaf-pS621 and activated Trigger signals; and a phenotype module which encoded the expression of functional differentiation markers from the ATRA-inducible transcription factors. The model described the transcription and translation of genes in each module, and signaling events in each module in a growing population of HL-60

cells. Model parameters were taken from literature, however, unknown coefficients that appear in the promoter logic models were estimated from protein measurements in HL-60 cells following ATRA exposure. Despite its simplicity, the effective model captured key features of the ATRA induced differentiation such as sustained MAPK activation, and bistability with respect to ATRA exposure. The model also described the expression of upstream transcription factors which regulated the expression of differentiation markers. Lastly, analysis of the response of the model to perturbations identified Gfi-1 and PPARγ as master regulators of ATRA-induced differentiation. We also found evidence of a prominent regulatory role for a signaling molecule ATRA-inducible component of the signalsome, Vav1. Vav1 is a guanine nucleotide exchange factor for Rho family GTPases that activate pathways leading to actin cytoskeletal rearrangements and transcriptional alterations [102]. The Vav1/Raf association correlated with Raf activity, was ATRA-inducible and decreased after treatment with the Raf inhibitor GW5074.

Naturally occurring cell fate decisions often incorporate reinforcing feedback and bistability [72, 142]. One of the most well studied cell fate circuits is the Mos mitogen-activated protein kinase cascade in *Xenopus* oocytes. This cascade is activated when oocytes are induced by the steroid hormone progesterone [249]. The MEK-dependent activation of p42 MAPK stimulates the accumulation of the Mos oncoprotein, which in turn activates MEK, thereby closing the feedback loop. This is similar to the signal initiation module presented here; ATRA drives signalsome formation, which activates MAPK, which in turn leads to more signalsome activation. Thus, while HL-60 and *Xenopus* oocytes are vastly different biological models, their cell fate programs share a similar architectural feature. Reinforcing feedback and bistability has also been implicated

in hematopoietic cell fate determination. Laslo et al showed in nonmalignant myelomonocytic cells that the counter antagonistic repressors, Gfi-1 and Egr-1/2 (whose expression is tuned by PU.1 and C/EBPα), encode a bistable switch that results in a macrophage, neutrophil or a mixed lineage population depending upon PU.1 and C/EBPα expression [142]. The current model contained the Gfi-1 and Egr-1/2 agonistic switch; however, its significance was unclear for HL-60 cells. The expression of Gfi-1, Egr-1/2, C/EBPα and PU.1 was not consistent with the canonical lineage pattern expected from literature. For example, Egr-1/2 expression (associated with a macrophage lineage) increased, while Gfi-1 expression (associated with a neutrophil lineage) was unchanged following ATRA exposure. Thus, HL-60 cells, which are a less mature cancer cell line, exhibited a non-canonical expression pattern. Other unrelated cell fate decisions such as programmed cell death have also been suggested to be bistable [5]. Still more biochemical networks important to human health, for example the human coagulation or complement cascades, also feature strong positive feedback elements [158]. Thus, while reinforcing feedback is often undesirable in human engineered systems, it is at the core of a diverse variety of cell fate programs and other networks important to human health.

Analysis of the signal integration and phenotype modules suggested Gfi-1 and PPARγ proteins were important regulators of ATRA-induced signal integration and phenotypic change. Model analysis showed that PU.1, Egr-1 and C/EBPα expression increased in Gfi-1$^{-/-}$ mutants, where PU.1 expression was upregulated by greater than 8-fold. Simulations suggested that combined Gfi-1 + PPARg deletion crippled the ability of HL-60 cells to undergo neutrophilic differentiation and function following ATRA exposure. This confirms previous literature showing that Gfi-1 KO mice lack normal neutrophils [98]. PU.1,

a member of the *ets* transcription factor family, is a well known regulator of granulocyte and monocyte development [76]. The relative level of PU.1 and C/EBPα is thought to control macrophage versus neutrophil cell fate decisions in granulocytic macrophage progenitor cells [43]. Simulations suggested that combined Gfi-1 + PPARγ deletion crippled the ability of HL-60 cells to undergo neutrophilic differentiation following ATRA exposure. p21 expression decreased significantly, suggesting Gfi-1$^{-/-}$ + PPAR$\gamma^{-/-}$ mutants were less likely to G0-arrest following ATRA exposure. The expression of other neutrophilic markers, such as CD38, also decreased in Gfi-1$^{-/-}$ + PPAR$\gamma^{-/-}$ cells. On the other hand, the expression of reactive oxygen metabolic markers, or other important transcription factors such as Oct4 did not change. For example, model analysis suggested that the C/EBPα dependent interaction of PU.1 with the *NCF1* gene, which encodes the p47Phox protein, was the most sensitive PU.1 connection; deletion of this connection removed the ability of the system to express p47Phox. p47Phox, also known as neutrophil cytosol factor 1, is one of four cytosolic subunits of the multi-protein NADPH oxidase complex found in neutrophils [63]. This enzyme is responsible for reactive oxygen species (ROS) production, a key component of the anti-microbial function of neutrophils. While p47Phox expression required C/EBPα and PU.1, neither Gfi-1 nor PPARγ deletion increased expression. This suggested that p47Phox expression was saturated with respect to C/EBPα and PU.1, and simultaneously not sensitive to PPARγ abundance. Taken together, Gfi-1$^{-/-}$ + PPAR$\gamma^{-/-}$ cells were predicted to exhibit some aspects of the ATRA response, but not other critical features such as cell cycle arrest. Hock et al showed that Gfi-1$^{-/-}$ mice lacked normal neutrophils, and were highly sensitive to bacterial infection [98]. Thus, the model analysis was consistent with this study. However, other predictions concerning

the behavior of the Gfi-$1^{-/-}$ + PPAR$\gamma^{-/-}$ mutants remain to be tested.

Immunoprecipitation studies identified a limited number of ATRA-dependent and -independent Raf interaction partners. We established potential interactions between Raf and key partners such as Vav1, Src, Akt, CK2 and 14-3-3. However, we were unable to detect the association of Raf with common kinases and GTPases such as PKC, PKA, p38, Rac and Rho as observed in literature [137, 57]. To investigate the association of c-Raf and PKA or PKC, we assessed both the expression levels and the associations of those important signaling molecules. Surprisingly, the expression levels of PKA, PKCα and PKCγ were not ATRA regulated, nor did we detect an association with Raf. We believe that our current perception of the signaling pathway driving differentiation in this model is novel. It also diverges from the classical perception in that its activation is not Ras driven. This has been reported by Katagiri et al that Ras is not a driver in RA-induced differentiation in HL-60 [127]. Finally, we note that the classical paradigms were typically derived in NIH3T3 cells where signal duration is rapid transient. By contrast in HL-60 it is a prolonged MAPK signaling that drives RA-induced granulocytic differentiation.

All of these partners are known to be associated with Raf activation or function. Src is known to bind Raf through an SH2 domain, and this association has been shown to be dependent of the serine phosphorylation of Raf [31]. Thus, an ATRA inducible Src/Raf association may be a result of ATRA-induced Raf phosphorylation at S259 or S621. We also identified an interaction between Raf and the Ser/Thr kinases Akt and CK2. Akt can phosphorylate Raf at S259, as demonstrated by studies in a human breast cancer line [264]. CK2 can also phosphorylate Raf, although the literature has traditionally focused on S338 and not

S621 or S259[193]. However, neither of these kinase interactions were ATRA-inducible, suggesting their association with Raf alone was not associated with ATRA-induced Raf phosphorylation. The adapter protein 14-3-3 was also constitutively associated with Raf. The interaction between Raf and 14-3-3 has been associated with both S621 and S259 phosphorylation and activity [91]. Additionally, the association of Raf with 14-3-3 not only stabilized S621 phosphorylation, but also reversed the S621 phosphorylation from inhibitory to activating [48]. Finally, we found that Vav1/Raf association correlated with Raf activity, was ATRA-inducible and decreased after treatment with GW5074. The presence of Vav1 in Raf/Grb2 complexes has been shown to correlate with increased Raf activity in mast cells [212]. Furthermore, studies on Vav1 knockout mice demonstrated that the loss of Vav1 resulted in deficiencies of ERK signaling for both T-cells as well as neutrophils [40, 84]. Interestingly, while an integrin ligand-induced ROS response was blocked in Vav1 knockout neutrophils, TPA was able to bypass the Vav1 requirement and stimulate both ERK phosphorylation and ROS induction [84]. It is possible that Vav1 is downstream of various integrin receptors but upstream of Raf in terms of inducible ROS responses. Vav1 has also been shown to associate with a Cbl-Slp76-CD38 complex in an ATRA-dependent manner; furthermore, transfection of HL-60 cells with Cbl mutants that fail to bind CD38, yet still bind Slp76 and Vav1, prevents ATRA-induced MAPK activation [208]. The literature suggest a variety of possible receptor-signaling pathways, which involve Vav1, for MAPK activation; moreover, given the ATRA-inducible association Vav1 may play a direct role in Raf activation.

We hypothesized that Vav1 is a member of an ATRA-inducible signalsome complex which propels sustained MAPK activation, arrest and differentiation (shown schematically in Fig. 2.9). Initially, ATRA-induced Vav1 expression

drives increased association between Vav1 and Raf. This increased interaction facilitates phosphorylation and activation of Raf by pre-bound Akt and/or CK2 at S621 or perhaps S259. Constitutively bound 14-3-3 may also stabilize the S621 phosphorylation, modulate the activity and/or up-regulate autophosphorylation. Activated Raf can then drive ERK activation, which in turn closes the positive feedback loop by activating Raf transcription factors e.g., Sp1 and/or STAT1 [135, 171, 260, 151]. We tested this working hypothesis using mathematical modeling. The model recapitulated both ATRA time-course data as well as the GW5074 inhibitor effects. This suggested the proposed Raf-Vav1 architecture was at least consistent with the experimental studies. Further, analysis of the Raf-Vav1 model identified bistability in phosphorylated ERK levels. Thus, two possible MAPK activation branches were possible for experimentally testable ATRA values. The analysis also suggested the ATRA-induced Raf-Vav1 architecture could be locked into a sustained signaling mode (high phosphorylated ERK) even in the absence of a ATRA signal. This locked-in property could give rise to an ATRA-induction memory. We validated the treatment memory property predicted by the Raf-Vav1 circuit experimentally using ATRA-washout experiments. ERK phosphorylation levels remained high for more then 96 hr after ATRA was removed. Previous studies demonstrated that HL-60 cells possessed an inheritable memory of ATRA stimulus [253]. Although the active state was self-sustaining, the inactive state demonstrated considerable robustness to perturbation. For example, we found that 50x overexpression of Raf was required to reliably lock MAPK into the activated state, while small perturbations had almost no effect on phosphorylated ERK levels over the entire ensemble. CD38 expression correlated with the phosphorylated ERK, suggesting its involvement in the signaling complex. Our computational and experimental results showed

that positive feedback, through ERK-dependent Raf expression, could sustain MAPK signaling through many division cycles. Such molecular mechanisms could underly aspects of cellular memory associated to consecutive ATRA treatments.

2.5 Methods

Effective gene expression model equations.

The ATRA differentiation model was encoded as a system of differential algebraic equations (DAEs) which described both signaling and gene expression processes. We modeled transcription and translation as Ordinary Differential Equations (ODEs), while signaling processes were assumed to quickly equilibrate and were treated as a pseudo steady state system of algebraic equations. The model formulation follows from a previous study of the Epithelial Messecnchymal Transition (EMT) [83]; in the current study additional attention was paid to the formulation of the transcription and translation rates, and an updated approach was taken to model the regulation of gene expression.

We decomposed the ATRA-induced differentiation program into three modules; a signal initiation module that sensed and transformed the ATRA signal into activated cRaf-pS621 and the ATRA-RAR/RXR (activated Trigger) signals; a signal integration module that controlled the expression of upstream transcription factors given cRaf-pS621 and activated Trigger signals; and a phenotype module which encoded the expression of functional differentiation markers from the ATRA-inducible transcription factors. The output of the signal

initiation module was the input to the gene expression model. For each gene $j = 1, 2, \ldots, \mathcal{G}$, we modeled both the mRNA (m_j), protein (p_j) and signaling species abundance:

$$\frac{dm_j}{dt} = r_{T,j} - \left(\mu + \theta_{m,j}\right) m_j + \lambda_j \tag{2.1}$$

$$\frac{dp_j}{dt} = r_{X,j} - \left(\mu + \theta_{p,j}\right) p_j \tag{2.2}$$

$$\mathbf{g}\left(p_1, \ldots, p_{\mathcal{G}}, \kappa\right) = \mathbf{0} \tag{2.3}$$

where signaling species abundance was governed by the non-linear algebraic equations $\mathbf{g}\left(p_1, \ldots, p_{\mathcal{G}}, \kappa\right) = \mathbf{0}$. The model parameter vector is denoted by κ. The terms $r_{T,j}$ and $r_{X,j}$ denote the specific rates of transcription, and translation while the terms $\theta_{m,j}$ and $\theta_{p,j}$ denote first-order degradation constants for mRNA and protein, respectively. The specific transcription rate $r_{T,j}$ was modeled as the product of a kinetic term $\bar{r}_{T,j}$ and a control term u_j which described how the abundance of transcription factors, or other regulators influenced the expression of gene j.

The gene expression control term $0 \leq u_j \leq 1$ depended upon the combination of factors which influenced the expression of gene j. If the expression of gene j was influenced by $1, \ldots, m$ factors, we modeled this relationship as $u_j = \mathcal{I}_j\left(f_{1j}(\cdot), \ldots, f_{mj}(\cdot)\right)$ where $0 \leq f_{ij}(\cdot) \leq 1$ denotes a regulatory transfer function quantifying the influence of factor i on the expression of gene j, and $\mathcal{I}_j(\cdot)$ denotes an integration rule which combines the individual regulatory inputs for gene j into a single control term. In this study, the integration rule governing gene expression was the weighted fraction of promoter configurations that

resulted in gene expression [174]:

$$u_j = \frac{W_{R_1,j} + \sum_n W_{nj} f_{nj}}{1 + W_{R_1,j} + \sum_d W_{dj} f_{dj}} \tag{2.4}$$

The numerator, the weighted sum (with weights W_{nj}) of promoter configurations leading to gene expression, was normalized by all possible promoter configurations (denominator). The likelihood of each configuration was quantified by the transfer function f_{nj} (which we modeled using Hill functions), while the lead term in the numerator $W_{R_1,j}$ denotes the weight of constitutive expression for gene j. Given the formulation of the control law, the λ_j term (which denotes the constitutive rate of expression of gene j) was given by:

$$\lambda_j = \bar{r}_{T,j} \left(\frac{W_{R_1,j}}{1 + W_{R_1,j}} \right) \tag{2.5}$$

The kinetic transcription term $\bar{r}_{T,j}$ was modeled as:

$$\bar{r}_{T,j} = V_T^{max} \left(\frac{L_{T,o}}{L_{T,j}} \right) \left(\frac{G_j}{K_T + G_j} \right) \tag{2.6}$$

where the maximum gene expression rate V_T^{max} was defined as the product of a characteristic transcription rate constant (k_T) and the abundance of RNA polymerase (R_1), $V_T^{max} = k_T (R_1)$. The $\left(L_{T,o}/L_{T,j} \right)$ term denotes the ratio of transcription read lengths; $L_{T,o}$ represents a characteristic gene length, while $L_{T,j}$ denotes the length of gene j. Thus, the ratio $\left(L_{T,o}/L_{T,j} \right)$ is a gene specific correction to the characteristic transcription rate V_T^{max}. If a gene expression process had no modi-

fying factors, $u_j = 1$. Lastly, the specific translation rate was modeled as:

$$r_{X,j} = V_X^{max}\left(\frac{L_{X,o}}{L_{X,j}}\right)\left(\frac{m_j}{K_X + m_j}\right) \tag{2.7}$$

where V_X^{max} denotes a characteristic maximum translation rate estimated from literature, and K_X denotes a translation saturation constant. The characteristic maximum translation rate was defined as the product of a characteristic translation rate constant (k_X) and the Ribosome abundance (R_2), $V_X^{max} = k_X(R_2)$. As was the case for transcription, we corrected the characteristic translation rate by the ratio of the length of a characteristic transcript normalized by the length of transcript j. The sequence lengths used in this study are given in Table 2.3; the characteristic gene and mRNA lengths were given by the average lengths computed from the values in Table 2.3.

Signaling model equations.

The signal initiation and integration modules required the abundance of cRaf-pS621 and ATRA-RAR/RXR (activated Trigger) as inputs. However, the base model described only the abundance of inactive proteins e.g., cRaf or RAR/RXR but not the activated forms. To address this issue, we estimated pseudo steady state approximations for the abundance of cRaf-pS621 and activated Trigger. The abundance of activated trigger ($x_{a,1}$) was estimated directly from the RAR/RXR abundance ($x_{u,1}$):

$$x_{a,1} \sim x_{u,1}\left(\frac{\alpha \cdot \text{ATRA}}{1 + \alpha \cdot \text{ATRA}}\right) \tag{2.8}$$

where α denotes a gain parameter; $\alpha = 0.0$ if ATRA is less than a threshold, and $\alpha = 0.1$ if ATRA is greater than the differentiation threshold. The abundance of cRaf-pS621 was estimated by making the pseudo steady state approximation on the cRaf-pS621 balance. In general, the abundance of an activated signaling species i was governed by:

$$\frac{dx_i}{dt} = r_{+,i}(\mathbf{x}, \mathbf{k}) - (\mu + k_{d,i})\, x_i \qquad i = 1, \ldots, \mathcal{M} \tag{2.9}$$

The quantity x_i denotes concentration of signaling species i, while $\mathcal{R}$ and $\mathcal{M}$ denote the number of signaling reactions and signaling species in the model, respectively. The term $r_{+,i}(\mathbf{x}, \mathbf{k})$ denotes the rate of generation of activated species i, while μ denotes the specific growth rate, and $k_{d,i}$ denotes the rate constant controlling the non-specific degradation of x_i. We neglected deactivation reactions e.g., phosphatase activities. We assumed that signaling processes were fast compared to gene expression; this allowed us to approximate the signaling balance as:

$$x_i^* \simeq \frac{r_{+,i}(\mathbf{x}, \mathbf{k})}{(\mu + k_{d,i})} \qquad i = 1, \ldots, \mathcal{M} \tag{2.10}$$

The generation rate was written as the product of a kinetic term ($\bar{r}_{+,i}$) and a control term (v_i). The control terms $0 \leq v_j \leq 1$ depended upon the combination of factors which influenced rate process j. If rate j was influenced by $1, \ldots, m$ factors, we modeled this relationship as $v_j = \mathcal{I}_j\left(f_{1j}(\cdot), \ldots, f_{mj}(\cdot)\right)$ where $0 \leq f_{ij}(\cdot) \leq 1$ denotes a regulatory transfer function quantifying the influence of factor i on rate j. The function $\mathcal{I}_j(\cdot)$ is an integration rule which maps the output of regulatory transfer functions into a control variable. In this study, we used $\mathcal{I}_j \in \{min, max\}$ and hill transfer functions [248]. If a process had no modifying

factors, $v_j = 1$. The kinetic rate of cRaf-pS621 generation $\bar{r}_{+,cRaf}$ was modeled as:

$$\bar{r}_{+,cRaf} = k_{+,cRaf} x_s \left(\frac{x_{cRaf}}{K_{+,cRaf} + x_{cRaf}} \right) \tag{2.11}$$

where x_s denotes the signalsome abundance, $k_{+,cRaf}$ denotes a characteristic activation rate constant, and $K_{+,cRaf}$ denotes a saturation constant governing cRaf-pS621 formation. In thus study, signalsome abundance was approximated by the abundance of the BLR1 protein; BLR1 expression is directly related to Raf nuclear translocation which in turn is related to activated signalsome. Thus, BLR1 is an indirect measure of the signalsome. The formation of cRaf-pS621 was regulated by only a single factor, the abundance of MAPK inhibitor, thus $v_{+,cRaf}$ took the form:

$$v_{+,cRaf} = \left(1 - \frac{\mathrm{I}}{K_D + \mathrm{I}} \right) \tag{2.12}$$

where I denotes the abundance of the MAPK inhibitor, and K_D denotes the inhibitor affinity.

Estimation of gene expression model parameters

Parameters appearing in the mRNA and protein balances, e.g., maximum transcription and translation rates, the half-life of a typical mRNA and proteins (assumed to be same for all transcripts/proteins), and typical values for the copies per cell of RNA polymerase and ribosomes were estimated from literature (Table 2.2). The saturation constants K_X and K_T appearing in the transcription and translation rate equations were adjusted so that gene expression and translation resulted in gene products on a biologically realistic concentration scale. Lastly, we calculated the concentration for gene G_j by assuming, on average, that a cell

had two copies of each gene at any given time. Thus, the bulk of our model parameters were taken from literature, and were not adjusted during model identification. However, the remaining parameters, e.g., the W_{ij} values or parameters appearing in the transfer functions f_{dj} which appeared in the gene expression control laws, were estimated from the experimental data discussed here. We assumed promoter configuration weights were bounded between $W_{ij} \in [0, 100]$; all cooperativity coefficients η_{ij} appearing in the binding transfer functions f_{dj} were bounded between $\eta_{ij} \in [0, 4]$; and all disassociation constants K_{ij} appearing in the binding transfer functions f_{dj} were bounded between $K_{ij} \in [0, 1000]$ (nM).

Signaling and gene expression model parameters were estimated by minimizing the squared difference between simulations and experimental protein data set j. We measured the squared difference in the scale, fold change and shape for protein j:

$$E_j(\mathbf{k}) = \left(\mathcal{M}_j(t_-) - \hat{y}_j(t_-, \mathbf{k})\right)^2 + \sum_{i=1}^{\mathcal{T}_j}\left(\hat{\mathcal{M}}_{ij} - \hat{y}_{ij}(\mathbf{k})\right)^2 + \sum_{i=1}^{\mathcal{T}_j}\left(\mathcal{M}'_{ij} - y'_{ij}(\mathbf{k})\right)^2 \tag{2.13}$$

The first term in Eq. (3.10) quantified the initial *scale* error, directly before the addition of ATRA. In this case, $\mathcal{M}_j(t_-)$ (the approximate concentration of protein j before the addition of ATRA) was estimated from literature. This term was required because the protein measurements were reported as the fold-change; thus, the data was normalized by a control value measured before the addition of ATRA. However, the model operated on a physical scale. The first term allowed the model to capture physically realistic changes following ATRA addition. The second term quantified the difference in the *fold-change* of protein j as a function of time. The terms $\hat{\mathcal{M}}_{ij}$ and $\hat{y}_{ij}$ denote the scaled experimental ob-

servations and simulation outputs (fold-change; protein normalized by control value directly before ATRA addition) at time i from protein j, where $\mathcal{T}_j$ denoted the number of time points for data set j. Lastly, the third term of the objective function measured the difference in the *shape* of the measured and simulated protein levels. The scaled value $0 \leq \mathcal{M}'_{ij} \leq 1$ was given by:

$$\hat{\mathcal{M}}_{ij} = \left(\mathcal{M}_{ij} - \min_i \mathcal{M}_{ij}\right) / \left(\max_i \mathcal{M}_{ij} - \min_i \mathcal{M}_{ij}\right) \tag{2.14}$$

where $\mathcal{M}'_{ij} = 0$ and $\mathcal{M}'_{ij} = 1$ describe the lowest (highest) intensity bands. A similar scaling was used for the simulation output. We minimized the total model residual $\sum_j E_j$ using a heuristic direct-search optimization procedure, subject to box constraints on the parameter values, starting from a random initial parameter guess. Each downhill step was archived and used for ensemble calculations. The optimization procedure (a covariance matrix adaptation evolution strategy) has been reported previously [110].

Estimation of an effective cell cycle arrest model.

We formulated an effective N-order polynomial model of the fraction of cells undergoing ATRA-induced cell cycle arrest at time t, $\hat{\mathcal{A}}(t)$, as:

$$\hat{\mathcal{A}}(t) \simeq a_0 + \sum_{i=1}^{N-1} a_i \phi_i \left(\mathbf{p}(t), t\right) \tag{2.15}$$

where a_i were unknown parameters, and $\phi_i \left(\mathbf{p}(t), t\right)$ denotes a basis function. The basis functions were dependent upon the system state; in this study, we used N

= 4 and basis functions of the form:

$$\phi_i\left(\mathbf{p}(t), t\right) = \left(\frac{t}{T} + \frac{p21}{E2F}\Big|_t\right)^{(i-1)} \tag{2.16}$$

The parameters $a_0, \ldots, a_3$ were estimated directly from cell-cycle measurements (biological replicates) using least-squares. The form of the basis function assumed p21 was directly proportional, and E2F inversely proportional, to G0-arrest. However, this was one of many possible forms for the basis functions.

Cell culture and treatment

Human myeloblastic leukemia cells (HL-60 cells) were grown in a humidified atmosphere of 5% CO_2 at 37^{o}C and maintained in RPMI 1640 from Gibco (Carlsbad, CA) supplemented with 5% heat inactivated fetal bovine serum from Hyclone (Logan, UT) and 1× antibiotic/antimicotic (Gibco, Carlsbad, CA). Cells were cultured in constant exponential growth [19]. Experimental cultures were initiated at 0.1×10^6 cells/mL 24 hr prior to ATRA treatment; if indicated, cells were also treated with GW5074 (2μM) 18 hr before ATRA treatment. For the cell culture washout experiments, cells were treated with ATRA for 24 hr, washed 3x with prewarmed serum supplemented culture medium to remove ATRA, and reseeded in ATRA-free media as described. Western blot analysis was performed at incremental time points after removal of ATRA.

Chemicals

All-Trans Retinoic Acid (ATRA) from Sigma-Aldrich (St. Louis, MO) was dissolved in 100% ethanol with a stock concentration of 5mM, and used at a final concentration of 1μM (unless otherwise noted). The cRaf inhibitor GW5074 from Sigma-Aldrich (St. Louis, MO) was dissolved in DMSO with a stock concentration of 10mM, and used at a final concentration of 2μM. HL-60 cells were treated with 2μM GW5074 with or without ATRA (1μM) at 0 hr. This GW5074 dosage had a negligible effect on the cell cycle distribution, compared to ATRA treatment alone.

Immunoprecipitation and western blotting

For immunoprecipitation experiments, cells were lysed as previously described. 300μg protein (in 300 μL total volume) per sample was pre-cleared with Protein A/G beads. The beads were pelleted and supernatant was incubated with Raf antibody (3μL/sample) and beads overnight. All incubations included protease and phosphatase inhibitors in M-PER used for lysis with constant rotation at 4^{o}C. Bead/antibody/protein slurries were then washed and subjected to standard SDS-PAGE analysis as previously described [208]. All antibodies were purchased from Cell Signaling (Boston, MA) with the exception of α-p621 Raf which was purchased from Biosource/Invitrogen (Carlsbad, CA), and α-CK2 from BD Biosciences (San Jose, CA).

Morphology assessment

Untreated and ATRA-treated HL-60 cells were collected after 72 hr and cyto-centrifuged for 3 min at 700 rpm onto glass slides. Slides were air-dried and stained with Wright's stain. Slide images were captured at 40X (Leica DM LB 100T microscope, Leica Microsystems).

Cell cycle quantification

1 x 10^6 cells were collected by centrifugation at 700 rpm and resuspended in 200 μl of cold propidium iodide (PI) hypotonic staining solution containing 50 μg/ml propidium iodine, 1 μl/ml Triton X-100, and 1 mg/ml sodium citrate (all Sigma, St. Louis, MO). Cells were incubated at room temperature for 1 h and analyzed by flow cytometry with a BD LSRII flow cytometer (BD Biosciences, San Jose, CA) using 488-nm excitation and emission collected through 550 long-pass dichroic and a 575/26 band-pass filters. Doublets were identified by a PI signal width versus area plot and excluded from the analysis.

CD11b quantification

Expression of cell surface differentiation markers was quantified by flow cytometry. 1 x 10^6 cells were collected from cultures and centrifuged at 700 rpm for 5 min. Cell pellets were resuspended in 200 μl 37°C PBS containing 2.5 μl of phycoerythrin (PE)-conjugated CD11b (both from BD Biosciences, San Jose, CA). Following 1 h incubation at 37°C cell surface expression levels were analyzed by flow cytometry. PE was excited at 488 nm and emission was collected through 505 long-pass dichroic and 530/30 band-pass filters.

Respiratory burst quantification (ROS)

1 x 10^6 cells were collected and centrifuged at 700 rpm for 5 min. Pellets were suspended in 500μl of PBS containing 5-(and-6)-chloromethyl-2',7'-dichlorodihydro-fluorescein diacetate acetyl ester (DCF, Molecular Probes, Eugene, OR) with either DMSO carrier blank solution or 12-0-tetradecanoyl-phorbol-13-acetate (TPA, Sigma, St. Louis, MO) suspended in DMSO. Cells were incubated for 25min at 37°C and then analyzed by flow cytometry as previously described. Oxidized DCF was excited by a 488nm laser and emission was collected through a 505 long pass dichroic mirror and a 530/30nm band pass filter. The shift in fluorescence intensity in response to TPA was used to determine the percent cells with the capability to generate inducible oxidative metabolites. Gates to determine percent positive cells were set to exclude 95% of control cells not stimulated with TPA.

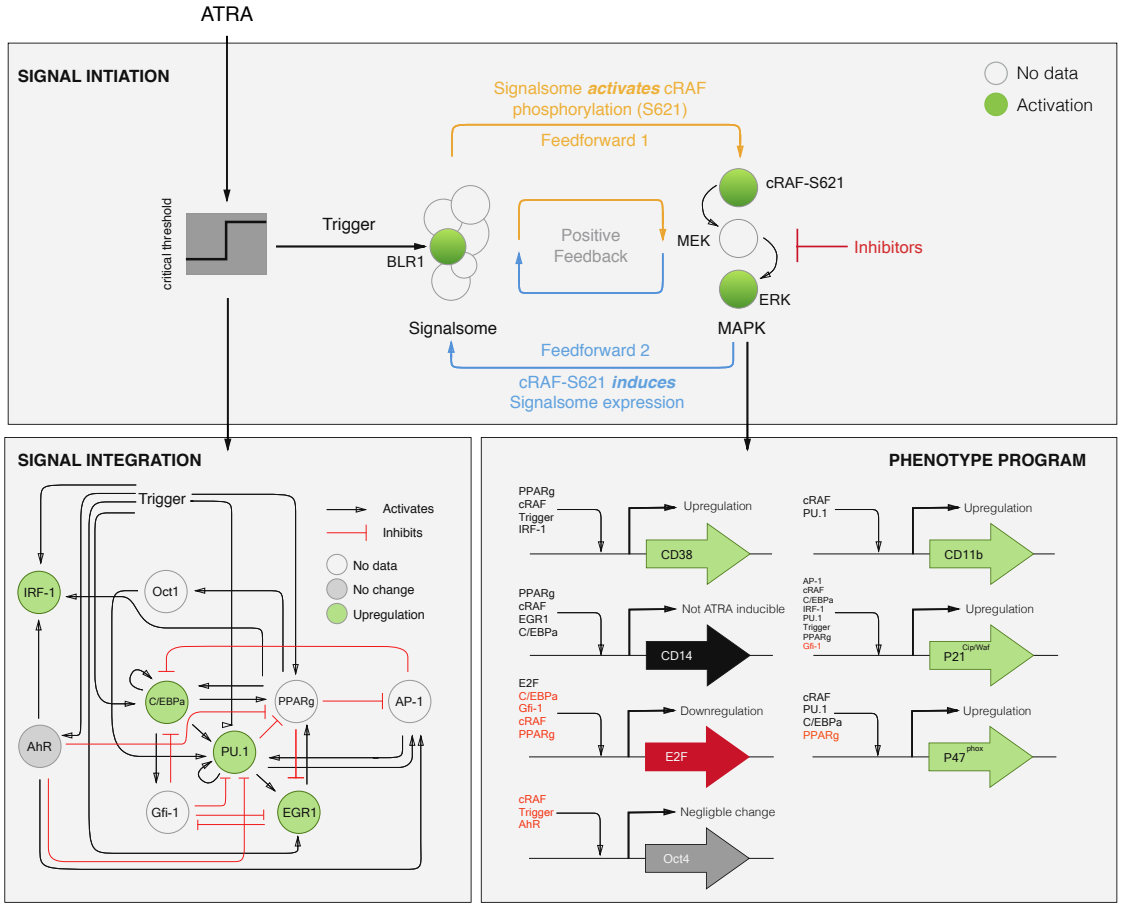

Figure 2.1: Schematic of the effective ATRA differentiation circuit. Above a critical threshold, ATRA activates an upstream Trigger, which induces signalsome complex formation. Signalsome activates the mitogen-activated protein kinase (MAPK) cascade which in turn drives the differentiation program and signalsome formation. Both Trigger and activated cRaf-pS621 drive a phenotype gene expression program responsible for differentiation. Trigger activates the expression of a series of transcription factors which in combination with cRaf-pS621 result in phenotypic change.

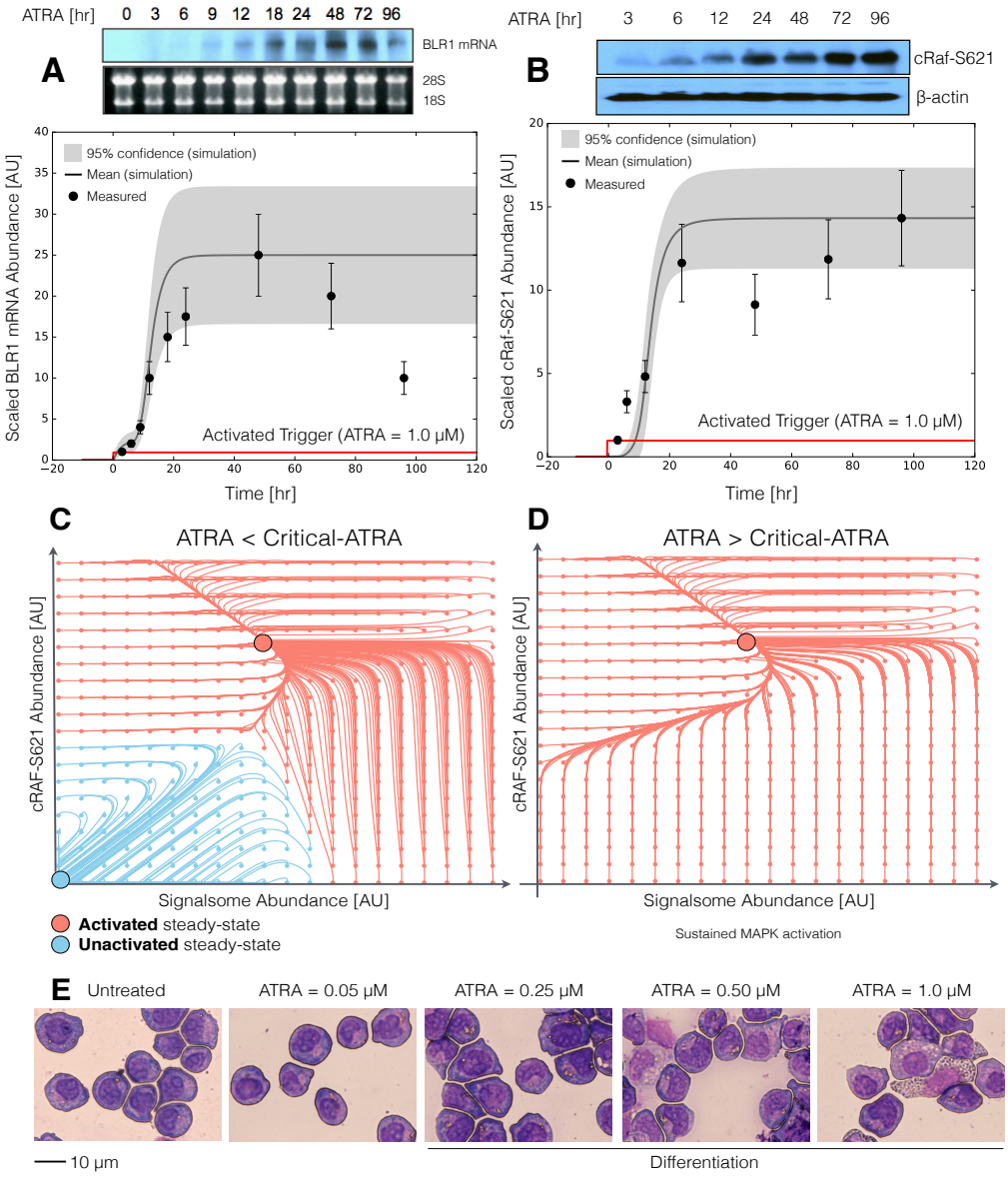

Figure 2.2: Model analysis for ATRA-induced HL-60 differentiation. A: BLR1 mRNA versus time following exposure to 1μM ATRA at t = 0 hr. B: cRaf-pS621 versus time following exposure to 1μM ATRA at t = 0 hr. Points denote experimental measurements, solid lines denote the mean model performance. Shaded regions denote the 99% confidence interval calculated over the parameter ensemble. C: Signalsome and cRaf-pS621 nullclines for ATRA below the critical threshold. The model had two stable steady states and a single unstable state in this regime. D: Signalsome and cRaf-pS621 nullclines for ATRA above the critical threshold. In this regime, the model had only a single stable steady state. E: Morphology of HL-60 as a function of ATRA concentration (t = 72 hr). Experimental data in panels A and B were reproduced from Wang and Yen [242], data in panel E is reported in this study.

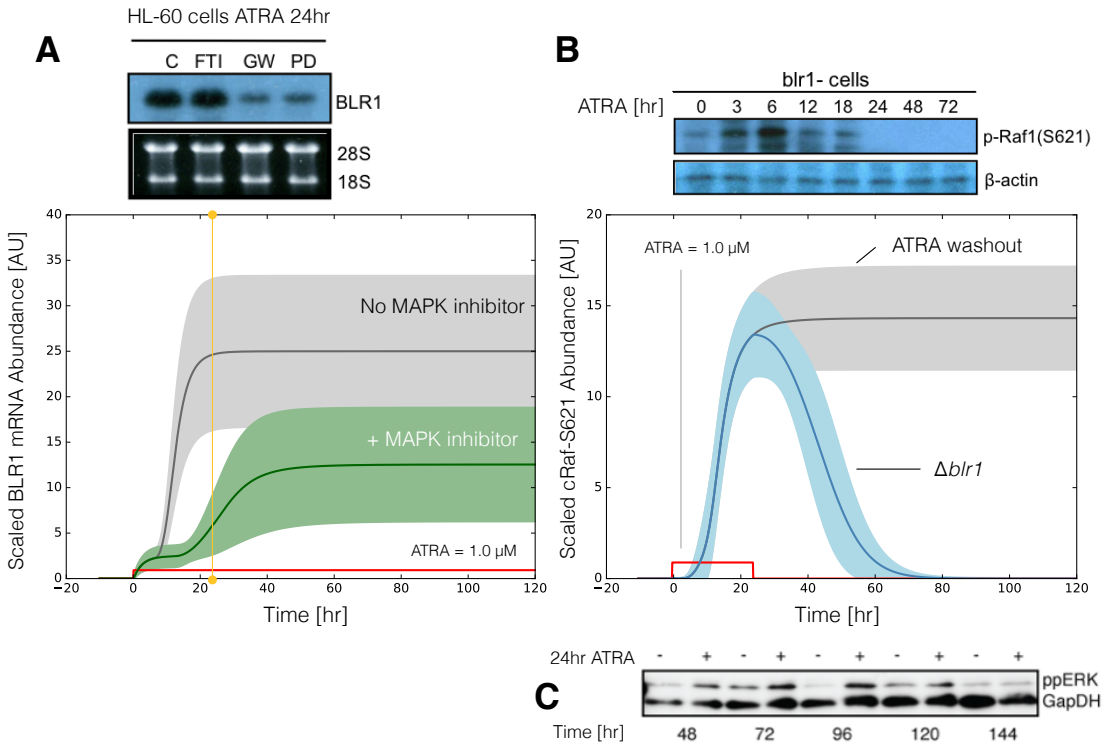

Figure 2.3: Model simulation following exposure to 1μM ATRA. A: BLR1 mRNA versus time with and without MAPK inhibitor. B: cRaf-pS621 versus time following pulsed exposure to 1μM ATRA with and without BLR1. Solid lines denote the mean model performance, while shaded regions denote the 99% confidence interval calculated over the parameter ensemble. C: Western blot analysis of phosphorylated ERK1/2 in ATRA washout experiments. Experimental data in panels A and B were reproduced from Wang and Yen [242], data in panel C is reported in this study. The images of the raw gel for panel C is given in the Supplemental Materials.

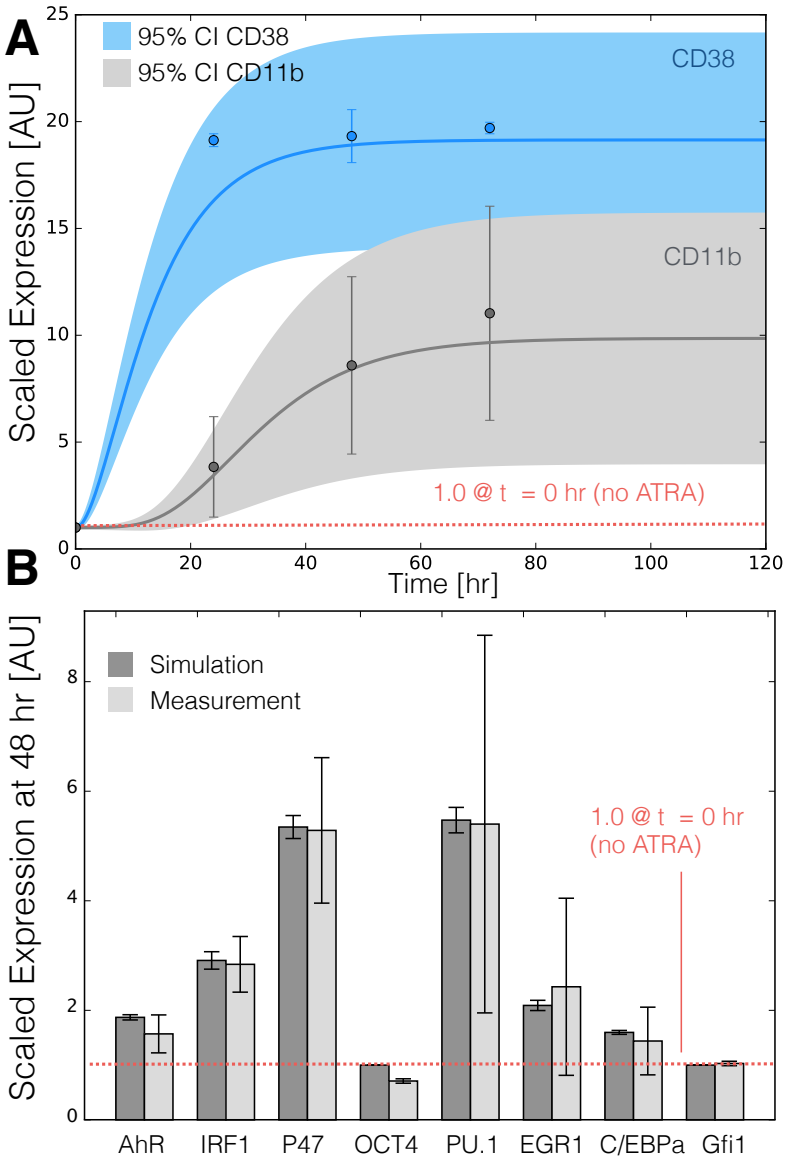

Figure 2.4: Model simulation of the HL-60 gene expression program following exposure to 1μM ATRA at t = 0 hr. A: Scaled CD38 and CD11b expression versus time following ATRA exposure at time t = 0 hr. B: Scaled Gene expression at t = 48 hr following ATRA exposure. Gene expression was normalized to expression in the absence of ATRA. The gene expression is quantified by the protein fold change of quantified Western blot data (from at least three biological repeat nuclear lysates) using ImageJ. Experimental data in panels A and B were reproduced from Jensen et al. [119]. Model simulations were conducted using the ten best parameter sets collected during model identification. Solid lines (or bars) denote the mean model performance, while the shaded region (or error bars) denote the 95% confidence interval calculated over the parameter ensemble.

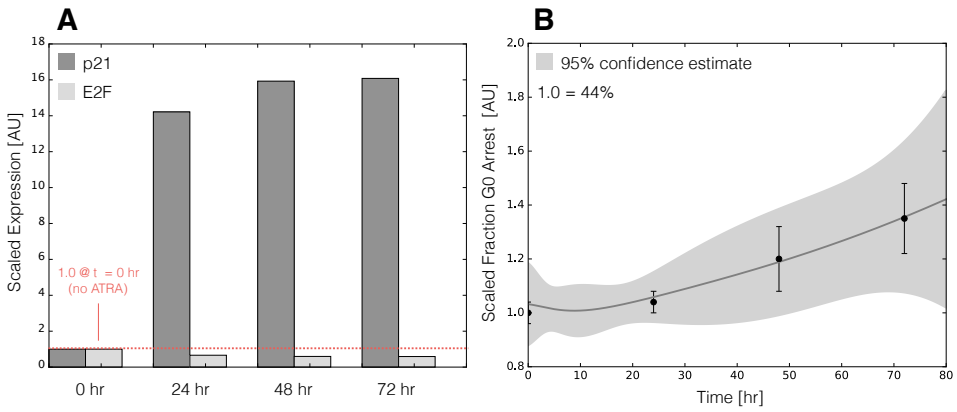

Figure 2.5: Model simulation of HL-60 cell-cycle arrest following exposure to 1μM ATRA at t = 0 hr. A: Predicted p21 and E2F expression levels for the best parameter set following ATRA exposure at time t = 0 hr. B: Estimated fraction of HL-60 cells in G0 arrest following ATRA exposure at time t = 0 hr. Solid lines (or bars) denote the mean model performance, while the shaded region (or error bars) denotes the 95% confidence estimate of the polynomial model. Experimental data in panel B was reproduced from Jensen et al. [119].

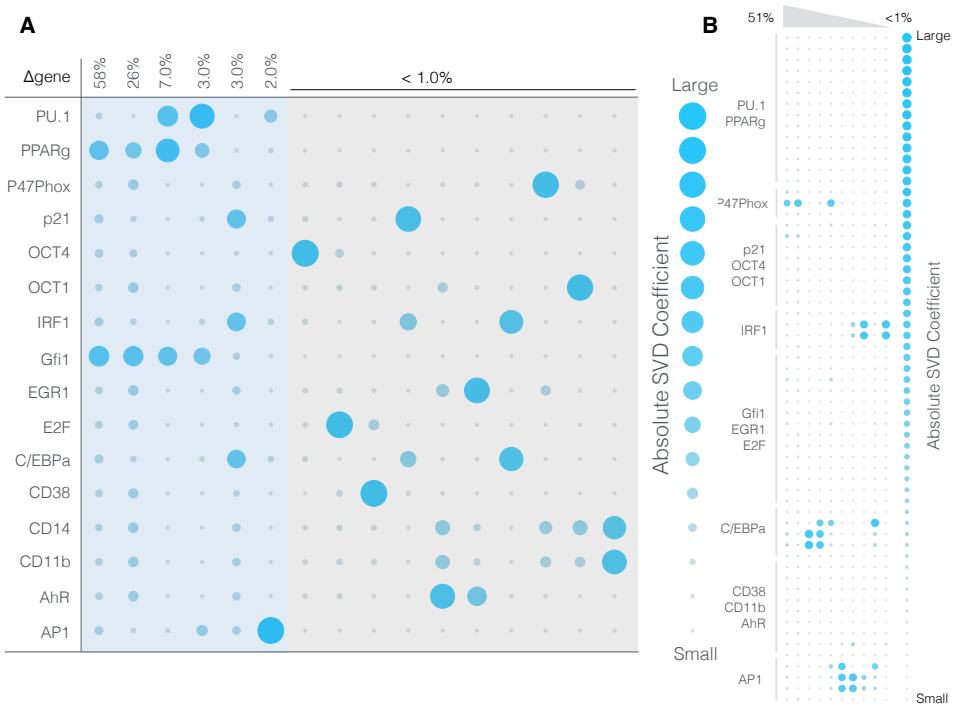

Figure 2.6: Robustness of the HL-60 differentiation program following exposure to 1μM ATRA at t = 0 hr. A: Singular value decomposition of the average system response (l^2-norm between the perturbed and nominal state) following pairwise gene knockout simulations using the top ten best fit parameter sets. The rows denote the deleted genes, while columns denote the response mode. B: Singular value decomposition of the average system response (l^2-norm between the perturbed and nominal state) following the pairwise removal of protein-DNA connections for the top ten best fit parameter sets. The rows denote protein-DNA interactions at the labeled promoter, while the columns denote the top ranked response modes (combinations of deletions). The percentage at the top of each column describes the fraction of the variance in the system state captured by the node combinations in the rows.

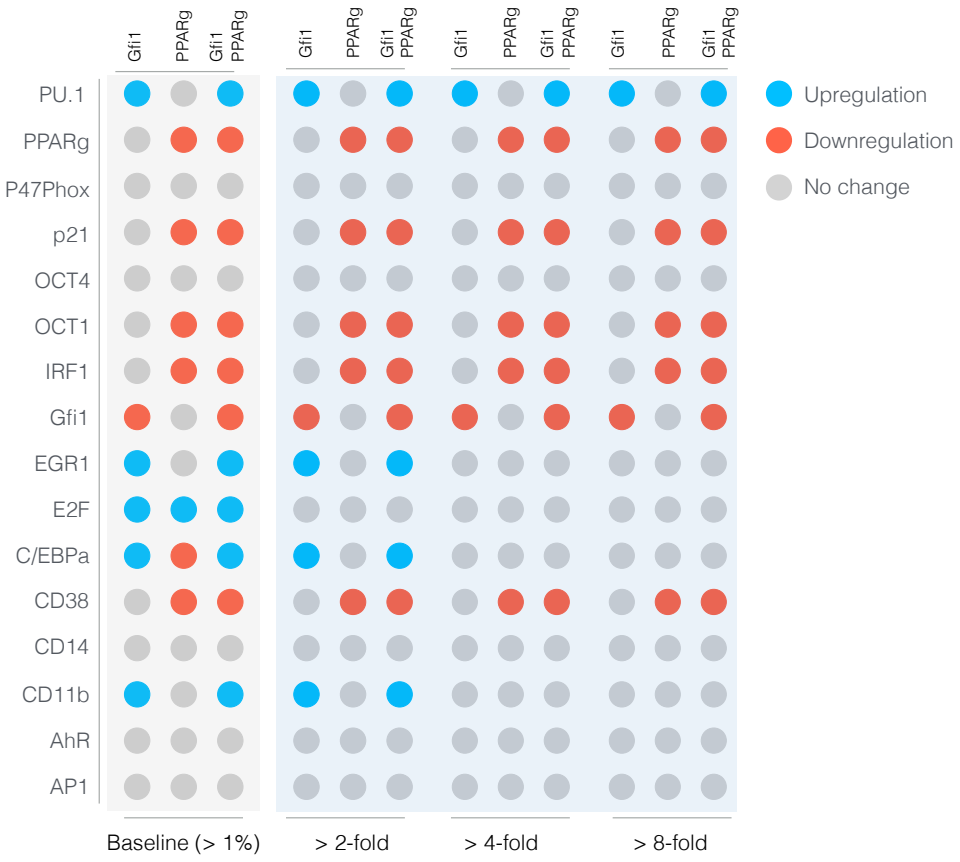

Figure 2.7: Robustness of the HL-60 differentiation program following exposure to 1μM ATRA at t = 0 hr. Protein fold change at t = 48 hr (rows) in single and double knock-out mutants (columns) relative to wild-type HL-60 cells. The responses were grouped into >2,4 and 8 fold changes. The best fit parameter set was used to calculate the protein fold change.

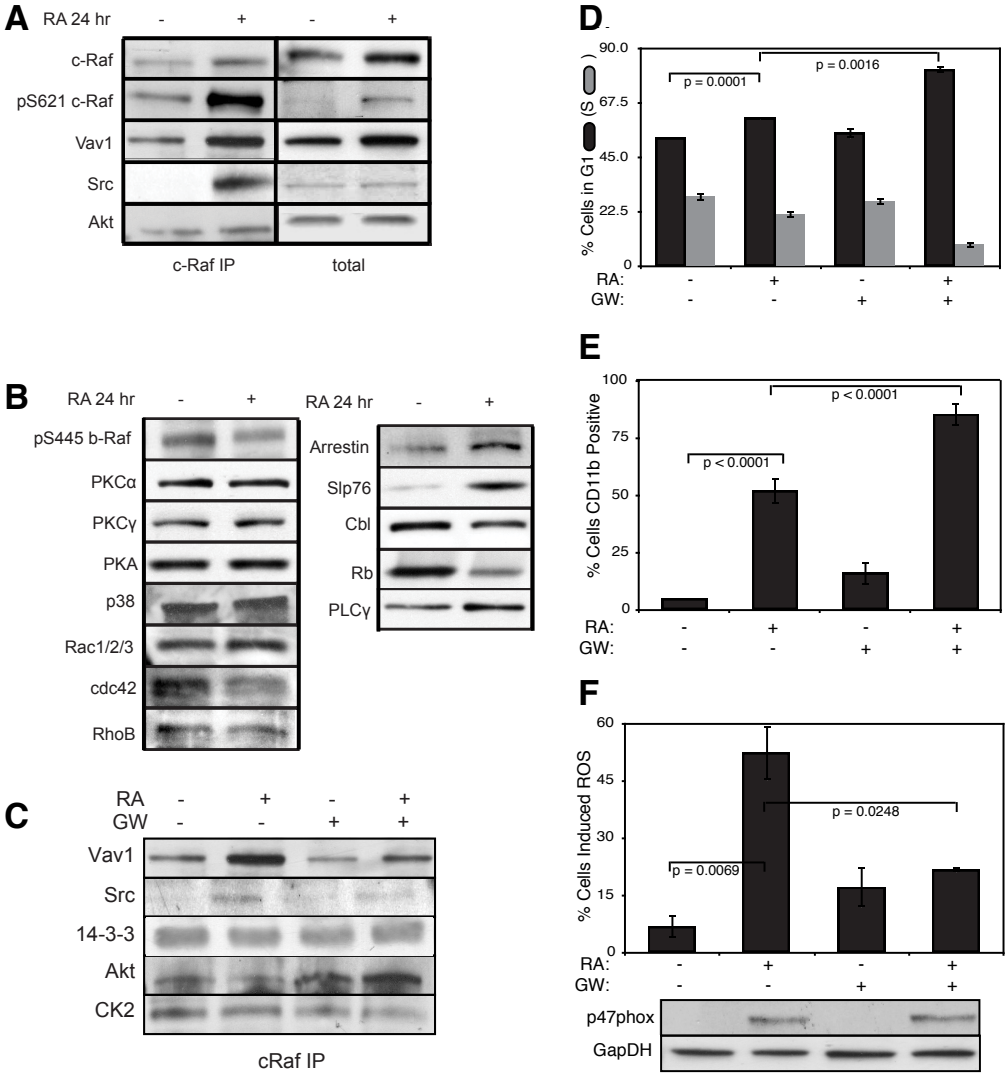

Figure 2.8: Investigation of a panel of possible Raf interaction partners in the presence and absence of ATRA. A: Species identified to precipitate out with Raf: first column shows Western blot analysis on total Raf immunoprecipitation with and without 24 hr ATRA treatment and the second on total lysate. B: The expression of species considered that did not precipitate out with Raf at levels detectable by Western blot analysis on total lysate. C: Effect of the Raf inhibitor GW5074 on Raf interactions as determined by Western blot analysis of total Raf immunoprecipitation. The Authors note the the signal associated with Src was weak. D: Cell Cycle distribution as determined by flow cytometry indicated arrest induced by ATRA, which was increased by the addition of GW5074. E: Expression of the cell surface marker CD11b as determined by flow cytometry indicated increased expression induced by ATRA, which was enhanced by the addition of GW5074. F: Inducible reactive oxygen species (ROS) as determined by DCF flow cytometry. The functional differentiation response of ATRA treated cells was mitigated by GW5074. GAPDH was used as a loading control. The black lines frame groupings from independent gels and each image

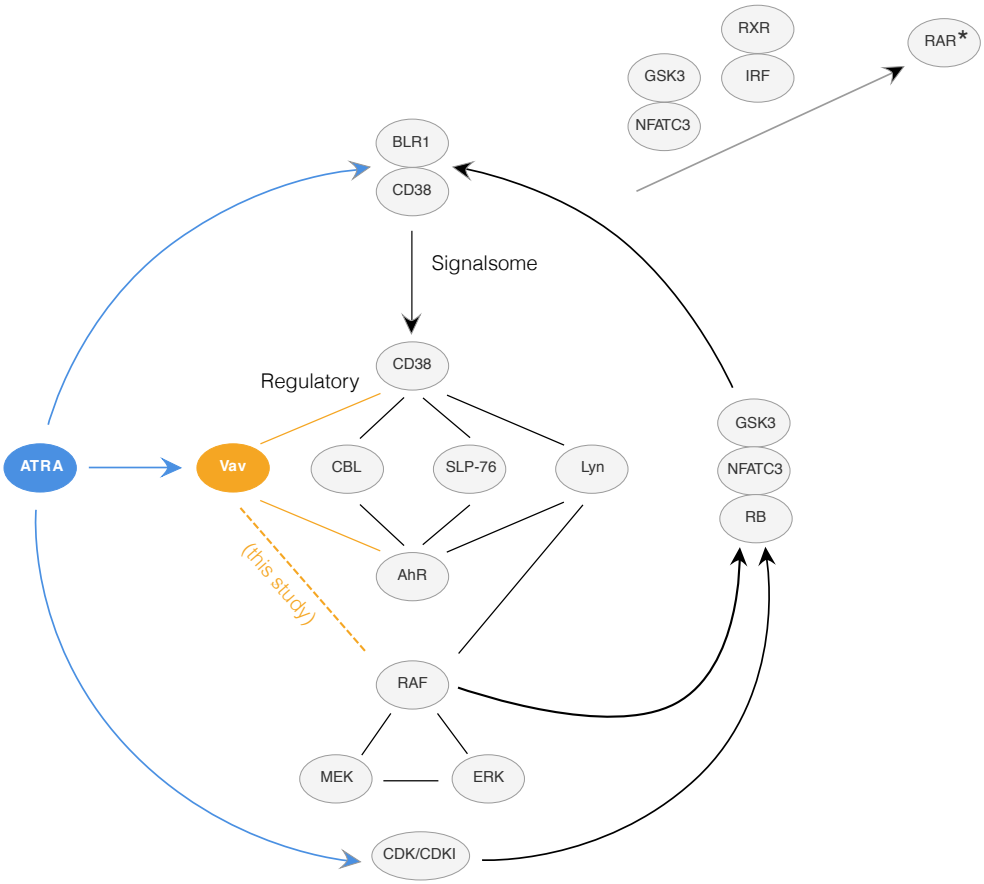

Figure 2.9: This schematic diagram shows the hypothetical principal pathways in the ATRA–induced signaling that results in cell differentiation in the HL-60 myeloid leukemia model [238, 34, 35, 208, 21, 207]. It is based on modules and feedback loops. There are three main arms (shown top to bottom): 1. Direct ATRA targeting of RAREs in genes such as CD38 or BLR1; 2. Formation of a signalsome that has a regulatory module that includes Vav (a guanine nucleotide exchange factor), CBL and SLP-76 (adaptors), and Lyn (a Src family kinase) that regulates a Raf/Mek/Erk axis that incorporates Erk to Raf feedback, where the regulators are modulated by AhR and CD38 receptors; and 3. Direct ATRA targeted up regulation of CDKI to control RB hypophosphorylation. The Raf/Mek/Erk axis is embedded in the signalsome and subject to modulation by the regulators. The output of the signalsome is discharge of the Raf from the cytosol to the nucleus where it binds (hyper)phospho-RB and other targets, including NFATc3, which enables activation of the ATRA bound RAR/RXR poised on the BLR1 promoter, and also GSK3, phosphorylation of which relieves its inhibitory effect on RARα. CDKI directed hypophosphorylation of RB releases Raf sequestered by RB to go to NFATc3, GSK3, and other targets. A significant consequence of the nuclear RAF is ergo ultimately to enable or hyperactivate transcriptional activation by RARα to drive differentiation. It might be noted that this proposed general model provides a mechanistic

Table 2.1: Myelomonocytic transcription factor connectivity used in the signal integration and phenotype modules.

Effector	Effect	Target	Source
RARα	+	RARα	[192]
	+	PU.1	[175]
	+	C/EBPα	[76]
	+	IRF-1	[160]
	-	Oct4	[220]
	+	CD38	[53]
	+	p21	[155]
	+	AhR	[23]
	+	Egr-1	[7]
PPARγ	+	C/EBPα	[195]
	+	IRF-1	[232]
	+	Oct1	[20]
	-	AP-1	[47]
	-	E2F	[4]
	-	Egr-1	[66]
	+	CD38	[211]
	+	CD14	[221]
	+	p21	[86]
	-	p47Phox	[236]
PU.1	-	PPARγ	[49]
	+	PU.1	[28]
	+	AP-1	[215]
	+	Egr-1	[142]
	+	CD11b	[185]
	+	p21	[257]
	+	p47Phox	[149]
C/EBPα	+	PPARγ	[195]

	+	PU.1	[43]
	+	C/EBPα	[227]
	+	Gfi-1	[153]
	-	E2F	[44]
	+	CD14	[187]
	+	p21	[89]
IRF-1	+	CD38	[9]
	+	p21	[189]
	-	PU.1	[42]
	-	C/EBPα	[54]
	-	E2F	[54]
	-	Egr-1	[142]
	-	p21	[54]
Oct1	+	PU.1	[29]
AP-1	-	PPARγ	[47]
	+	PU.1	[11]
	+	p21	[125]
E2F	+	E2F	[123]
Egr-1	+	PPARγ	[77]
	-	Gfi-1	[163]
	+	CD14	[27]
AhR	+	AP-1	[217]
	+	IRF-1	[206]
	-	Oct4	[22]
	-	PU.1	[14]

Table 2.2: Characteristic model parameters estimated from literature.

Symbol	Description	Value	Units	Source
R_1	RNA polymerase abundance	85,000	copies/cell	[112, 261]
R_2	Ribosome abundance	1 x 10^6	copies/cell	[75]
G_i	Characteristic gene abundance	2	copies/cell	this study
K_X	Saturation constant transcription	600	copies/cell	this study
K_T	Saturation constant translation	95,000	copies/cell	this study
$t_{1/2,m}$	characteristic mRNA half-life (transcription factor)	2-4	hr	[250]
$t_{1/2,p}$	characteristic protein half-life	10	hr	[50]
$\theta_{m,j}$	characteristic mRNA degradation constant	0.34	hr^{-1}	derived
$\theta_{p,j}$	characteristic protein degradation constant	0.07	hr^{-1}	derived
t_d	HL-60 doubling time	19.5	hr	this study
μ	growth rate	0.035	hr^{-1}	derived
k_d	death rate	0.10μ	hr^{-1}	derived
e_T	elongation rate RNA polymerase	50-100	nt/s	[202, 46, 140, 45]
e_X	elongation rate Ribosome	5	aa/s	[15]
$L_{T,o}$	characteristic gene length	44,192	nt	[169]
$L_{X,o}$	characteristic transcript length	1,374	nt	derived
k_T	characteristic transcription rate	1.44	hr^{-1}	derived
k_X	characteristic translation rate	3.60	hr^{-1}	derived
$k_{+,cRaf}$	characteristic cRaf-S621 activation rate constant	1	hr^{-1}	this study
$K_{+,cRaf}$	characteristic saturation constant governing cRaf-pS621 formation	60	nM	this study
K_D	characteristic MAPK inhibitor affinity	10	nM	this study
D	Diameter of an HL-60 cell	12.4	μm^3	[196]
f_C	cytoplasmic fraction	0.51	dimensionless	[196]

Table 2.3: Sequence lengths from NCBI RefSeq database were used in the signal integration and phenotype modules [184]. The RNA sequence length used represents the total distance of transcription, and assume to be equal to the gene length.

Gene Name	Gene (bp)	RNA (bp)	Protein (aa)	Gene ID	Protein ID
AP-1	10323	10323	331	Gene ID: 3725	NP_002219
AhR	47530	47530	848	Gene ID: 196	NP_001621
CD11b	72925	72925	1153	Gene ID: 3684	NP_001139280
CD14	8974	8974	375	Gene ID: 929	NP_001035110
CD38	174978	74978	300	Gene ID: 952	NP_001766
C/EBPα	2630	2630	393	Gene ID: 1050	NP_001274353.1
E2F	17919	17919	437	Gene ID: 1869	NP_005216
Egr-1	10824	10824	543	Gene ID: 1958	NP_001955
Gfi-1	13833	13833	422	Gene ID: 2672	NP_005254
IRF-1	16165	16165	325	Gene ID: 3659	NP_002189
Oct1	206516	206516	741.33	Gene ID: 5451	NP_002688.3, NP_001185712.1, NP_001185715.1
Oct4	6356	6356	206.33	Gene ID: 5460	NP_001167002, NP_001167015, NP_001167016
P21	15651	15651	198	NG_009364.1	NP_001621
P47	3074	3074	390	GenBank: AF003533.1	NP_000256
PPARγ	153507	153507	250	Gene ID: 5468	NP_001317544
PU.1	40782	40782	270.5	Gene ID: 6688	NP_001074016, NP_003111

CHAPTER 3

MODELING AND ANALYSIS OF COMBINATION TREATMENT OF HL60 TRANSCRIPTION FACTOR NETWORK

3.1 Abstract

In this study, we present an effective model All-Trans Retinoic Acid (ATRA)-induced differentiation of HL-60 cells. The model describes a key architectural feature of ATRA-induced differentiation, positive feedback between an ATRA-inducible signalsome complex involving many proteins including Vav1, a guanine nucleotide exchange factor, and the activation of the mitogen activated protein kinase (MAPK) cascade. The model, which was developed by integrating logical rules with kinetic modeling, was significantly smaller than previous models. However, despite its simplicity, it captured key features of ATRA induced differentiation of HL-60 cells. We identified an ensemble of effective model parameters using measurements taken from ATRA-induced HL-60 cells. Using these parameters, model analysis predicted that MAPK activation was bistable as a function of ATRA exposure. Conformational experiments supported ATRA-induced bistability. These findings, combined with other literature evidence, suggest that positive feedback is central to a diversity of cell fate programs.

Keywords: leukemia, cancer, mathematical modeling

3.2 Introduction

It was suggested by Bailey, more than a decade ago, that qualitative and quantitative knowledge of complex biological systems could be achieved in the absence of complete structural and parameter knowledge [**?**]. Although this is true, the incomplete knowledge of biological phenomenon often limits the impact of computational models. Unknown or even disputed network structures can lead to incomplete fitting of computation models, requiring additional experiments and updating of the model. Since ODE kinetic models typically require extensive prior knowledge of network structure, rate constants and initial conditions [**?**], often a single "correct" network structure is assumed. Villaverde *et al.* discussed three main strategies in the reverse engineering of dynamic models: (1) full network inference, (2) network selections, and (3) kinetic parameter estimations [**?**]. Strategy 1, where the kinetic model structure and kinetic parameters are unknown, is typically solved by identifying the model interaction network without dynamics and then identifying the kinetic parameters. In this study, we will consider strategy 2; we have an initial model network structure that will be perturbed to find modifications to fit the experimental data for six different leukemia cell lines.

Leukemia is the 6th leading cause of cancer death in both males and females in the United States [**?**]. Approximately 72% of cancer related leukemia deaths were caused by four main types of leukemia; acute lymphocyctic leukemia (ALL), chronic lymphocytic leukemia (CLL), acute myeloid leukemia (AML), and chronic myeloid leukemia (CML). ALL and CLL are characterized by accumulation of lymphocytes in the bone marrow, with ALL progressing at a faster rate (i.e. acute vs. chronic). ALL occurs in both children and adults and has

a 90% five year survival rate in children [**?**], while CLL is rare in children and has a 66% five year survival rate (2001-2009) [**?**]. The 10 year survival rate for CML has improved to 80-90% due to the use of targeted treatments for BCR-ABL and adenosine triphosphate (ATP) [**?**]. AML is characterized by the accumulation of abnormally differentiated cells of the hematopoietic system in the bone marrow and blood, with a survival rate of 35 to 40% in adults under the age of 60 (5 to 15% in patients older than 60) [**?**]. AML is a group of extremely heterogeneous diseases, with over 200 known chromosome translocations and mutations in patient leukemic cells [**?**]. The use of differentiation induction therapy agents like all-*trans* retinoic acid (RA) and 1,25-dihyrodyvitamin D3 (D3) have been explored in many cancer cell types, including myeloid leukemias, and lung, liver, prostate, and breast cancer (RA treatment) [**?**, **?**] and in prostate, breast, colorectal, leukemia, and brain (D3 treatment) [**?**]. Acute promyelocytic leukemia (APL), a subtype of AML, was once one of the most fatal forms of acute leukemia until the introduction of RA increased remission rates of patients to between 80 and 90 % [**?**]. Failure of RA treatment can occur initially in patients with RA resistant variants (PLZF-RARA-positive APL), and relapse occurs in 5-20% of cases due to the emergence of RA resistance. To understand the response of multiple leukemia cell lines to RA and D3 treatments, we developed a network structure of well known transcription factors governing myelomonocytic lineage selection (granulocytic or monocytic).

In this study we considered data from RA, D3, and RA plus D3 treatments on six human myeloid leukemia cell lines; (1) K562 (FAB M1), (2) HL60 (FAB M2), (3) NB4 (FAB M3), (4) U937 (FAB 5), (5) HL60 R38+ and (6) HL60 R38- (5 and 6 are described previously [**?**]). K562, a CML cell line with a Bcr-Abl fusion protein [**?**], was used as a control because the cells are not responsive to

both RA [**?**] and D3 treatment [**?**]. HL60 cells, an AML cell line, are lineage-bipotent myelobasts [**?**, **?**] that can differentiate to either granulocytic lineage (using RA) or monocytic lineage (using D3). The two RA-resistant HL60 cell lines, R38+ and R38-, were described previously [**?**, **?**]. NB4, an APL cell line, are highly RA-responsive, but require combination treatments for monocytic differentiation (i.e. low response to D3) [**?**, **?**]. U937, histiocytic lymphoma cell line, are highly responsive to D3 induced monocytic differentiation, but have ambiguous differentiation effects due to RA (either monocytic or granulocytic) [**?**, **?**, **?**]. The model network contained two inputs (RA and D3) and 18 main species. We included two receptors, reinoic acid receptor alpha (RARα) and vitamin D receptor (VDR), but excluded their heterodimer receptor pair, retinoid receptor (RXR), for simplicity (assumption RXR is readily available). Also, included in the model were transcription factors important in myelomonocytic lineage selection (listed in Table **??**). Finally upstream markers for differentiation were included, including CD38, CD11b, CD14, G1/GO cell cycle arrest and inducible oxidative metabolism. CD38 and CD11b are myelomonocytic markers and CD14 is a monocytic specific marker.

In this study we first determined that a network structure of a small three node protein model could be identified with sufficient experimental data. A model structure search in combination with particle swarm optimization to determine parameter values narrowed down total possible model structures from 19683 to twenty. After additional experimental data was implemented, we were able to find the synthetic model structure with no *a priori* knowledge. Next we investigated possible network structures for transcription factor and upstream markers in six leukemia cell lines treated with RA and D3 (data from [**?**]). The upstream markers CD38 and CD11b (myelomonocytic markers) and

CD14 (monocytic specific marker) were included in the model. Also, included in the model were transcription factors important in myelomonocytic lineage selection. Starting from an initial best model structure curated from literature sources, we were able to improve the model fits for six leukemia cell lines versus experimental data.

3.3 Results

We constructed an effective model of ATRA and D3 induced HL-60 differentiation which described gene expression events following the addition of ATRA and D3 (Fig. 3.1). The model connectivity was developed from literature and the studies presented here (Table 3.1). In this study, we present the ATRA and D3 program into two modules; a signal integration module that controlled the expression of upstream transcription factors given cRaf-pS621 and activated Trigger signals (Fig. 3.1); and a phenotype module which encoded the expression of functional differentiation markers from the ATRA-inducible transcription factors (Fig. 5.1). Each component of these modules was described by a mRNA and protein balance equation. Lastly, because the population of HL-60 cells was dividing (at least before ATRA-induced cell cycle arrest), we also considered a dilution term in all balance equations. The signal integration and phenotype modules were collectively encoded by 54 differential equations. Model parameters were taken literature (Table 2.2), or estimated from experimental data using heuristic optimization (see materials and methods).

The signal integration module that controlled the expression of upstream transcription factors takes signaling cues from an upstream module (not in-

cluded) as the input. The signal initiation module produced two outputs, activated Triggers (ATRA and D3) and cRaf-pS621 which drove the expression of ATRA and D3-induced transcription factors. The expression of transcriptions factors initiated a cascade gene expression that activated the phenotypic program. In particular, Trigger ATRA, which is a surrogate for the RARα/RXR transcriptional complex, regulated the expression of the transcription factors CCATT/enhancer binding protein α (C/EBPα), PU.1, and EGR1. Meanwhile, Trigger D3 is a surrogate for the VDR/RXR transcriptional complex that regulates the expression of PPARγ and PU.1. In turn, these transcription factors, in combination with cRaf-pS621, regulated the expression of downstream phenotypic markers such as CD38, CD11b or P47Phox.

We assembled the connectivity of the signal integration and phenotypic programs driven by the Triggers and cRaf-pS621 from literature (Table 3.1). The signaling events driven by ATRA activates the ATRA trigger that drives the transcription activation (TA) of IRF-1, AhR, PPARg, and master regulators CEBPa and PU.1. Analogously, D3 activates D3 trigger and drives the TA of PPARg and PU.1. These transcription factors directly induced by the activation of ATRA and D3 triggers in turn drives or inhibits expression of other gene in the gene regulatory network. A combination of these expression levels and the sustained activation of MAPK are then used to determine the expression of phenotypic markers in the phenotypical module.

To make sense of such a complicated network with multiple incoherent interaction, we used effective kinetics coupled with logical rules that describes transcription regulation. We estimated the parameters which appeared in the control laws regulating these programs from steady-state and dynamic mea-

surements of transcription factor and phenotypic marker expression following the addition of ATRA and/or D3 [REFHERE]. However, the bulk of the remaining model parameters were taken from directly from literature [172] and were not estimated in this study (see materials and methods) which created reduced the number of undetermined parameters. The model simulations captured the time dependent expression of CD38, CD11b, and CD14 following the addition of different combinations of ATRA and D3 treatments.

The induction of ATRA induces expression of surface markers that are associated with myeloid differentiation, CD38 and CD11b but not CD14. Using our model, we generated an ensemble set of 495 parameter sets that was able to capture the transcritpion factor expression measured at 48 hours in the signal integration module (Fig 3.2 and the time course expression of key phenotypic markers (Fig. 3.3). Although we capture most of the TF expression at 48 hours, the model fail to capture some combination treatment results showed synergistic or antonisitic effect when compared to the sum of the ATRA and D3 treatments. For example, EGR1 and GFI1 shoedw synergistic effect, while AhR showed antagonistic effects. For the time course simulations, the black lines represent the ensemble means, the dark colored bands are the 99% standard error from the mean, and the light colored bands are the 99% confidence interval. Similarly, the induction of D3 induces expression of surface markers that are associated with monocyte differentiation, CD38, CD11b, and CD14 are all expressed. Using the network structure, and 491 parameter ensemble, we were able to capture the time dependent phenotypic markers(Fig. 3.3). Finally, we tested to see if our model was able to capture the complexity involved with differentiation induced with ATRA and D3 at the same time (Fig. 3.3).Taken together, the signal integration and phenotypic simulations were consistent with

measurements, thereby validating the assumed molecular connectivity.

A question we wanted to answer was how ATRA induced differentiation differs from D3 or combination induced differentiation , and how might the role of parameters change based on the different treatments. To help answer the question, we performed sensitivity analysis on the model and the parameter sensitivity for the treatments were ranked from low to high. The sensitivity rankings of the treatments were plotted against one another for direct comparison (ATRA vs D3, ATRA vs combination, and D3 vs combination; Figs. 3.4-3.6) with respect to 9 species. The most sensitive parameters have higher rank order, while the lest sensitive parameters are localized towards the origin. The analysis showed the weight parameters (W; that govern the individual contribution to gene expression associated with individual TF binding) to be the most sensitive across all treatments. Parameters that diverged from the 45 degree axis indicated a change in parameter sensitivity between the who treatments compared. The labels on the sensitivity figure represent the parameter associated with a gene and influences of transcription factors. A closer analysis was taken at the measurable phenotypical markers of CD11b, CD14, CD38, and p47Phox. The parameters that were sensitive to change between treatments for CD11b were related to the regulation of PU1 across all comparisons, with the highest change in sensitivity (distance from 45 degree axis) in the saturation constants. Interestingly, the parameters sensitivity associated with the expression of CD11b did not change between treatments. Next we looked at CD14, the monocytic surface marker that is only induced by D3. Since CD14 is not ATRA inducible, sensitivity comparison of ATRA alone with respect to CD14 showed a high number of sensitivity changes. However when treatments of D3 and combination were compared (Fig. 3.6, the comparison showed that with D3 alone, the expres-

sion of CD14 relies more on the expression of PPARγ, while the combination of ATRA and D3 treatment showed higher sensitivity for EGR1 and CEBPα. CD38 is an important early marker in both granucyctic and monocytic differentiation. Although PPARγ is a key TF that regulates CD38 expression, ATRA induced differentiation is more sensitive to the induction of PPARγ through EGR1 while D3 showed higher sensitivity towards PPARγ expression through PU1 activation. Interestingly, when ATRA was compared to combination treatment (Fig. 3.5) for CD38, only the weight of PU1 on the expression of PPARγ showed a change in sensitivity. However CD38 sensitivity for D3 compared to combination showed a variety of changes such as OCT1 basal expression having a higher sensitivity for combination treatment. P47phox, the NADPH oxidase required for phagocytes to generate reactive oxygen species (ROS) for host defense, was used as functional markers for differentiation. We compared the P47phox parameter sensitivity between ATRA induced and D3 induced differentiation and combination treatments. While P47phox relies on the TF binding of PU1 and CEBPα, the analysis showed that PU1 is sensitive for ATRA-induced differentiation, while CEBPα is more sensitive for D3-induced differentiation.

To make sense of the change in parameter sensitivities across the different treatments, we counted the off-axis frequencies of each parameters on all the protein species and presented them in Fig. 3.7. The grey points are the comparison between ATRA and D3, blue represents the comparison between ATRA and combination, and red is the comparison etween D3 and combination. The positive frequencies indicate the parameter favored the respective treatment represented above zero and vise versa. The parameters with the highest absolute frequencies indicates their change in sensitivity across all species between the treatments compared. When ATRA and D3 induced differentiation were com-

pared, the parameter sensitivities that heavily favored ATRA treatment were related to Trigger ATRA induced gene expression (w parameters of Trigger ATRA activating expressions of PPARγ, EGR1, CEBPa, and PU1). Interestingly, parameter that showed up off-axis the most for ATRA vs D3 was the weight parameters of Trigger ATRA's regulation of PPARγ that was most sensitive to ATRA treatment, however when D3 was combined with ATRA, the parameter gained even more sensitivity compared to ATRA alone (Fig. 3.7 blue). Lastly as a validation step of our model, we plotted the cost function of our ensemble of parameters against two 495-parameter-set ensembles that was randomly and semi-randomly generated using sobol sequences (Fig. 3.8. The error plot showed that the best fit parameter ensemble performed better across all cost functions and that performance was not solely due to the structure of the model.

3.4 Discussion

In this study, we presented an effective model of ATRA and D3 inducible differentiation of HL-60 cells which integrates signals into a gene regulatory network and expressed measurable phenotypic markers. Despite its simplicity, the model captured key features of the ATRA and D3 induced differentiation such as the expressions of CD38, CD11b, and CD14.

Sensitivity analysis showed that the parameters sensitivity associated with differential marker expression were robust across the different treatments, but rather the change in sensitivity comes from parameters associated with TF expression that impact the downstream phenotypic markers. This finding may seem trivial but may offer some insight into responders and non-responders

of ATRA treatment. Performing this series of sensitivity analysis on ATRA-resistant cells that can be rescued with combination treatment shown by Jensen and coworkers (REFERENCE) will reveal a great amount of detail in how the sensitivity changes upon rescue. Although we capture most of the TF expression at 48 hours, the model fail to capture some combination treatment results showed synergistic or antonisitic effect when compared to the sum of the ATRA and D3 treatments. For example, EGR1 and GFI1 showed synergistic effect, while AhR showed antagonistic effects. Within the network, the expression of AhR is activated by Trigger ATRA, therefore the level was increased during ATRA-induced differentiation and having no change during D3-induced differentiation. However, in the presence of both ATRA and D3, AhR showed a decrease in concentration compared to the baseline at time zero although there is no TF that inhibit it's expression in the network. The balance of AhR concentration is determined by the rate of production and the rate of degradation. AhR degradation is ubiquitin mediated and occurs via the 26S proteasome pathway (REFERENCE: https://www.ncbi.nlm.nih.gov/pubmed/12213384). The activation of both Trigger ATRA and D3 may increase AhR ubiquination leading to a higher rate of degradation, we look to further explore this mechanism in the future. Similarly, we captured the 48hr expression of PU1 for monotherapy of ATRA and D3-induced differentiation, however the combination treatment yielded interesting results because the addition of D3 to ATRA treatment reduced the 48hr expression for PU1. This nonaddictive effect may be due to the upregulation of GFI1 at 48hrs from the combination treatment that was observed in the experiments but not capture in our simulations. GFI1 inhibits the expression of PU1, therefore the higher level GFI1 for the combination case will contribute to more inhibited level of PU1 expression, leading to a lower

level at 48 hours. This brings up the questions of why GFI1 level is synergistic increased with combination treatment, while not increased from monotreatments for ATRA or D3 alone. The expression of GFI1 is induced by one TF, CEBPα, which is not heavily upregulated in the combination treatment compared to monotreatments, therefore the upregulation of GFI1 must be to addition biology that we do not account for in our model. Similarly, EGR1 also behaves in a similar manner. Both GFI1 and EGR1 and TFs that determines the terminal fate of differentiation with GFI1 and EGR1 supporting granulcytic and monotypic differentiation respectively. The combination treatment show synergistic effect of expression of the two terminal regulators, and the mechanism behind the synergistic effect is still undetermined. The upregulation of GFI1 is not only involved in hemopoesis, but Xing and coworkers also showed that GFI1 downregulation promotes inflammation-linked metastasis of colorectal cancer (https://www.nature.com/articles/cdd201750). Analogously, EGR1 is a direct regulator of multiple tumor suppressors including TGFβ1, PTEN, p53 and fibronectin (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2455793/). There is a huge opportunity in understanding the mechanisms behind the synergistic effect of ATRA and D3 that overexpresses GFI1 and EGR1 for the treatment of resistance subpopulation or other types of cancer.

CD38 is an important early marker in both granucyctic and monocytic differentiation. Although PPARγ is a key TF that regulates CD38 expression, ATRA induced differentiation is more sensitive to the induction of PPARγ through EGR1 while D3 showed higher sensitivity towards PPARγ expression through PU1 activation. This finding suggest that parameters directly associated with CD38 expression are robust and the sensitivtiy does not change with respect to different treatment, therefore any effect that changes the expression of CD38

would be due to upstream TFs that regulate CD38 rather than CD38 regulation itself. This hypothesis is further supported by the sensitivity comparison between ATRA and D3 induced differentiation where EGR1 and PU1 were the key players in PPARγ and CD38 expression for ATRA and D3 treatments respectively. The two monotreatment took different pathways to activate and express the terminal phenotypic marker. With a better quantitative understanding of the signal integration circuitry and the new analytical methodology, we believe this is the first step in understanding ATRA resistance. With model aided interrogation, we can identify which portion of the network that may be unresponsive or identify unexplained mechanisms in the unresponsive cases.

3.5 Materials and Methods

Gene expression model equations.

We decomposed the ATRA-induced differentiation program into three modules; a signal initiation module that sensed and transformed the ATRA signal into activated cRaf-pS621 and the ATRA-RXR/RAR (activated Trigger) signals; a signal integration module that controlled the expression of upstream transcription factors given cRaf-pS621 and activated Trigger signals; and a phenotype module which encoded the expression of functional differentiation markers from the ATRA-inducible transcription factors. The output of the signal initiation module was the input to the gene expression model. For each gene $j = 1, 2, \ldots, \mathcal{G}$,

we modeled both the mRNA (m_j) and protein (p_j) abundance:

$$\frac{dm_j}{dt} = r_{T,j} - \left(\mu + \theta_{m,j}\right) m_j + \lambda_j \tag{3.1}$$

$$\frac{dp_j}{dt} = r_{X,j} - \left(\mu + \theta_{p,j}\right) p_j \tag{3.2}$$

The terms $r_{T,j}$ and $r_{X,j}$ denote the specific rates of transcription, and translation while the terms $\theta_{m,j}$ and $\theta_{p,j}$ denote first-order degradation constants for mRNA and protein, respectively. The specific transcription rate $r_{T,j}$ was modeled as the product of a kinetic term $\bar{r}_{T,j}$ and a control term u_j which described how the abundance of transcription factors, or other regulators influenced the expression of gene j. The kinetic transcription term $\bar{r}_{T,j}$ was modeled as:

$$\bar{r}_{T,j} = V_T^{max} \left(\frac{L_{T,o}}{L_{T,j}}\right)\left(\frac{G_j}{K_T + G_j}\right) \tag{3.3}$$

where the maximum gene expression rate V_T^{max} was defined as the product of a characteristic transcription rate constant (k_T) and the abundance of RNA polymerase (R_1), $V_T^{max} = k_T\,(R_1)$. The $\left(L_{T,o}/L_{T,j}\right)$ term denotes the ratio of transcription read lengths; $L_{T,o}$ represents a characteristic gene length, while $L_{T,j}$ denotes the length of gene j. Thus, the ratio $\left(L_{T,o}/L_{T,j}\right)$ is a gene specific correction to the characteristic transcription rate V_T^{max}. The degradation rate constants were defined as $\theta_{m,j}$ and $\theta_{p,j}$ denote characteristic degradation constants for mRNA and protein, respectively. Lastly, the λ_j term denotes the constitutive rate of expression of gene j.

The gene expression control term $0 \leq u_j \leq 1$ depended upon the combination of factors which influenced the expression of gene j. If the expression of gene j was influenced by $1, \ldots, m$ factors, we modeled this relationship as

$u_j = \mathcal{I}_j\left(f_{1j}(\cdot), \ldots, f_{mj}(\cdot)\right)$ where $0 \leq f_{ij}(\cdot) \leq 1$ denotes a regulatory transfer function quantifying the influence of factor i on the expression of gene j, and $\mathcal{I}_j(\cdot)$ denotes an integration rule which combines the individual regulatory inputs for gene j into a single control term. In this study, the integration rule governing gene expression was the weighted fraction of promoter configurations that resulted in gene expression [REFHERE]:

$$u_j = \frac{W_{R_1,j} + \sum_n W_{nj} f_{nj}}{1 + W_{R_1,j} + \sum_d W_{dj} f_{dj}} \tag{3.4}$$

The numerator, the weighted sum (with weights W_{nj}) of promoter configurations leading to gene expression, was normalized by all possible promoter configurations. The likelihood of each configuration was quantified by the transfer function f_{nj} (which we modeled using hill like functions), while the lead term in the numerator $W_{R_1,j}$ denotes the weight of constitutive expression for gene j. Given this formulation, the rate of constitutive expression was then given by:

$$\lambda_j = \bar{r}_{T,j}\left(\frac{W_{R_1,j}}{1 + W_{R_1,j}}\right) \tag{3.5}$$

If a gene expression process had no modifying factors, $u_j = 1$. Lastly, the specific translation rate was modeled as:

$$r_{X,j} = V_X^{max}\left(\frac{L_{X,o}}{L_{X,j}}\right)\left(\frac{m_j}{K_X + m_j}\right) \tag{3.6}$$

where V_X^{max} denotes a characteristic maximum translation rate estimated from literature, and K_X denotes a translation saturation constant. The characteristic

maximum translation rate was defined as the product of a characteristic translation rate constant (k_X) and the Ribosome abundance (R_2), $V_X^{max} = k_X\,(R_2)$. As was the case for transcription, we corrected the characteristic translation rate by the ratio of the length of a characteristic transcription normalized by the length of transcript j.

Effective cell cycle arrest model.

A defining feature of ATRA-induced differentiation is G1/G0 cell cycle arrest followed by differentiation. While we included E2F transcription, we do not consider a detailed model of restriction point regulation. Instead, we formulated an effective model of the fraction of cells undergoing ATRA-induced cell cycle arrest, $\mathcal{A}$ as:

$$\frac{d\mathcal{A}}{dt} = \epsilon_1\,(\mathrm{E2F})^{\epsilon_2} \tag{3.7}$$

The parameters ϵ_1 and ϵ_2 were estimated from experimental cell distribution measurements by `log` transforming Eq. (3.7):

$$\log\epsilon_1 + \epsilon_2 \log(\mathrm{E2F}_1) = \log\dot{\mathcal{A}}_{21} \tag{3.8}$$

$$\log\epsilon_1 + \epsilon_2 \log(\mathrm{E2F}_2) = \log\dot{\mathcal{A}}_{32} \tag{3.9}$$

where subscripts denote the experimental time point (1 = 24 hr, 2 = 48 hr, and 3 = 72 hr), $\mathrm{E2F}_i$ denotes the E2F abundance at time point i, and $\dot{\mathcal{A}}_{ij}$ denotes the time derivative of the fraction of arrested cells (estimated from data). We calculated candidate values for ϵ_j by solving the log-transformed system of equations for 10 biological replicates.

Estimation of model parameters.

We estimated parameters appearing in the mRNA and protein balances, and the abundance of polymerases and ribosomes, from estimates of transcription and translation rates, the half-life of a typical mRNA and protein, and typical values for the copies per cell of RNA polymerase and ribosomes from literature (Table ZZ). For the remaining parameters, e.g., the W_{ij} appearing in the control laws, or parameters appearing in the transfer functions f_{dj}, were estimated from the gene expression and signaling data sets discussed here. The saturation constants K_X and K_T were adjusted so that gene expression and translation resulted in gene products on a biologically realistic concentration scale. Lastly, we calculated the concentration for gene G_j by assuming, on average, that a cell had two copies of each gene at any given time. Thus, the bulk of our gene expression parameters were based directly upon literature values, and were not adjusted during model identification. The values used for the characteristic transcription/translation parameters, degradation constants and macromolecular copy number are given in the supplemental materials along with the specific formulas required to calculate all derived constants.

Signal and gene expression model parameters were estimated by minimizing the squared difference between simulations and experimental data set j:

$$E_j(\mathbf{k}) = \sum_{i=1}^{\mathcal{T}_j} \left(\hat{\mathcal{M}}_{ij} - \hat{y}_{ij}(\mathbf{k}) \right)^2 + \left(\frac{\mathcal{M}'_{ij} - \max y_{ij}}{\mathcal{M}'_{ij}} \right)^2 \tag{3.10}$$

The terms $\hat{\mathcal{M}}_{ij}$ and $\hat{y}_{ij}$ denote scaled experimental observations and simulation outputs at time i from training set j, where $\mathcal{T}_j$ denoted the number of time points for data set j. The first term in Eqn. (3.10) quantified the relative simulation

error. We used immunoblot intensity measurements for model training. Thus, we trained the model on the *relative* change between bands within each data set. Suppose we have the intensity of species x at time $\{t_1, t_2, .., t_n\}$ in condition j. The scaled value $0 \leq \hat{\mathcal{M}}_{ij} \leq 1$ is given by:

$$\hat{\mathcal{M}}_{ij} = \left(\mathcal{M}_{ij} - \min_i \mathcal{M}_{ij}\right) / \left(\max_i \mathcal{M}_{ij} - \min_i \mathcal{M}_{ij}\right) \tag{3.11}$$

where $\hat{\mathcal{M}}_{ij} = 0$ and $\hat{\mathcal{M}}_{ij} = 1$ describe the lowest (highest) intensity bands. A similar scaling was used for the simulation output. The second term in the objective function ensured a realistic concentration scale was estimated by the model. We set the highest intensity band to $\mathcal{M}'_{ij} = 10$ [AU] for all simulations. We minimized the total model residual $\sum_j E_j$ using heuristic optimization starting from a random initial parameter guess.

The signaling and gene expression model equations were implemented in the Julia programming language, and solved using the `ODE23s` routine of the ODE package [12]. The model code and parameter ensemble is freely available under an MIT software license and can be downloaded from http://www.varnerlab.org. Model parameters estimated directly or derived from literature, or data presented in this study are given in Table 2.2.

Cell culture and treatment

Human myeloblastic leukemia cells (HL-60 cells) were grown in a humidified atmosphere of 5% CO_2 at 37^{o}C and maintained in RPMI 1640 from Gibco (Carlsbad, CA) supplemented with 5% heat inactivated fetal bovine serum from Hyclone (Logan, UT) and 1× antibiotic/antimicotic (Gibco, Carlsbad, CA). Cells

were cultured in constant exponential growth [19]. Experimental cultures were initiated at 0.1×10^6 cells/mL 24 hr prior to ATRA treatment; if indicated, cells were also treated with GW5074 (2μM) 18 hr before ATRA treatment. For the cell culture washout experiments, cells were treated with ATRA for 24 hr, washed 3x with prewarmed serum supplemented culture medium to remove ATRA, and reseeded in ATRA-free media as described. Western blot analysis was performed at incremental time points after removal of ATRA.

Chemicals

All-Trans Retinoic Acid (ATRA) from Sigma-Aldrich (St. Louis, MO) was dissolved in 100% ethanol with a stock concentration of 5mM, and used at a final concentration of 1μM (unless otherwise noted). The cRaf inhibitor GW5074 from Sigma-Aldrich (St. Louis, MO) was dissolved in DMSO with a stock concentration of 10mM, and used at a final concentration of 2μM. HL-60 cells were treated with 2μM GW5074 with or without ATRA (1μM) at 0 hr. This GW5074 dosage had a negligible effect on the cell cycle distribution, compared to ATRA treatment alone.

Immunoprecipitation and western blotting

Approximately 1.2×10^7 cells were lysed using 400μL of M-Per lysis buffer from Thermo Scientific (Waltham, MA). Lysates were cleared by centrifugation at 16,950 $\times$ g in a micro-centrifuge for 20 min at 4^{o}C. Lysates were pre-cleared using 100μL protein A/G Plus agarose beads from Santa Cruz Biotechnology (Santa Cruz, CA) by inverting overnight at 4^{o}C. Beads were cleared by centrifu-

gation and total protein concentration was determined by a BCA assay (Thermo Scientific, Waltham, MA). Immunoprecipitations were setup by bringing lysate to a concentration of 1g/L in a total volume of 300μL (M-Per buffer was used for dilution). The anti-Raf antibody was added at 3μL. A negative control with no bait protein was also used to exclude the direct interaction of proteins with the A/G beads. After 1 hr of inversion at 4^{o}C, 20μL of agarose beads was added and samples were left to invert overnight at 4^{o}C. Samples were then washed three times with M-Per buffer by centrifugation. Finally proteins were eluted from agarose beads using a laemmli loading buffer. Eluted proteins were resolved by SDS-PAGE and Western blotting. Total lysate samples were normalized by total protein concentration (20μg per sample) and resolved by SDS-PAGE and Western blotting. Secondary HRP bound antibody was used for visualization. All antibodies were purchased from Cell Signaling (Boston, MA) with the exception of α-p621 Raf which was purchased from Biosource/Invitrogen (Carlsbad, CA), and α-CK2 from BD Biosciences (San Jose, CA).

Morphology assessment

Untreated and ATRA-treated HL-60 cells were collected after 72 hr and cytocentrifuged for 3 min at 700 rpm onto glass slides. Slides were air-dried and stained with Wrightâs stain. Slide images were captured at 40X (Leica DM LB 100T microscope, Leica Microsystems).

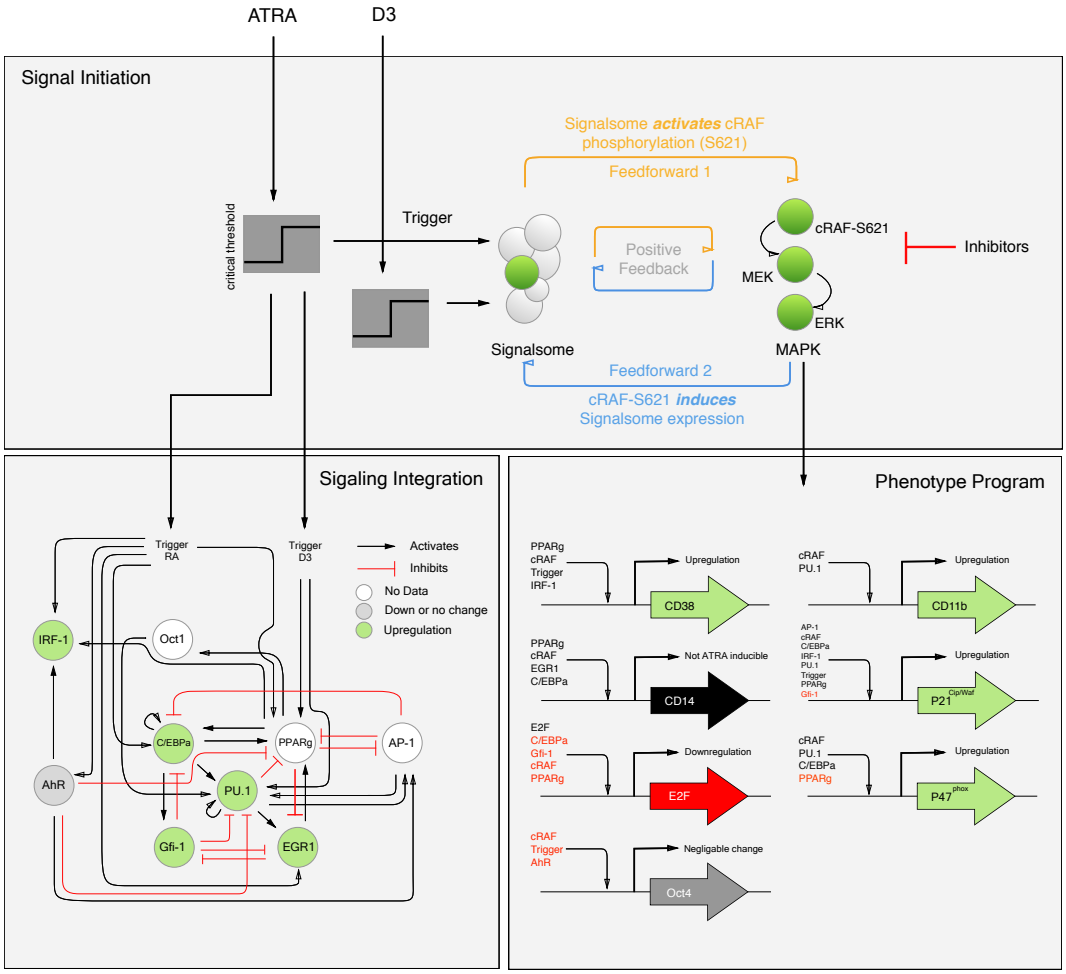

Figure 3.1: HL60 transcription factor network structure from literature sources. Figure shows interactions between protein nodes with black arrows indicating transcriptional activation and red lines indicating transcriptional repression. Phenotypic markers are shown on the right, with black indicating activation and red repression.

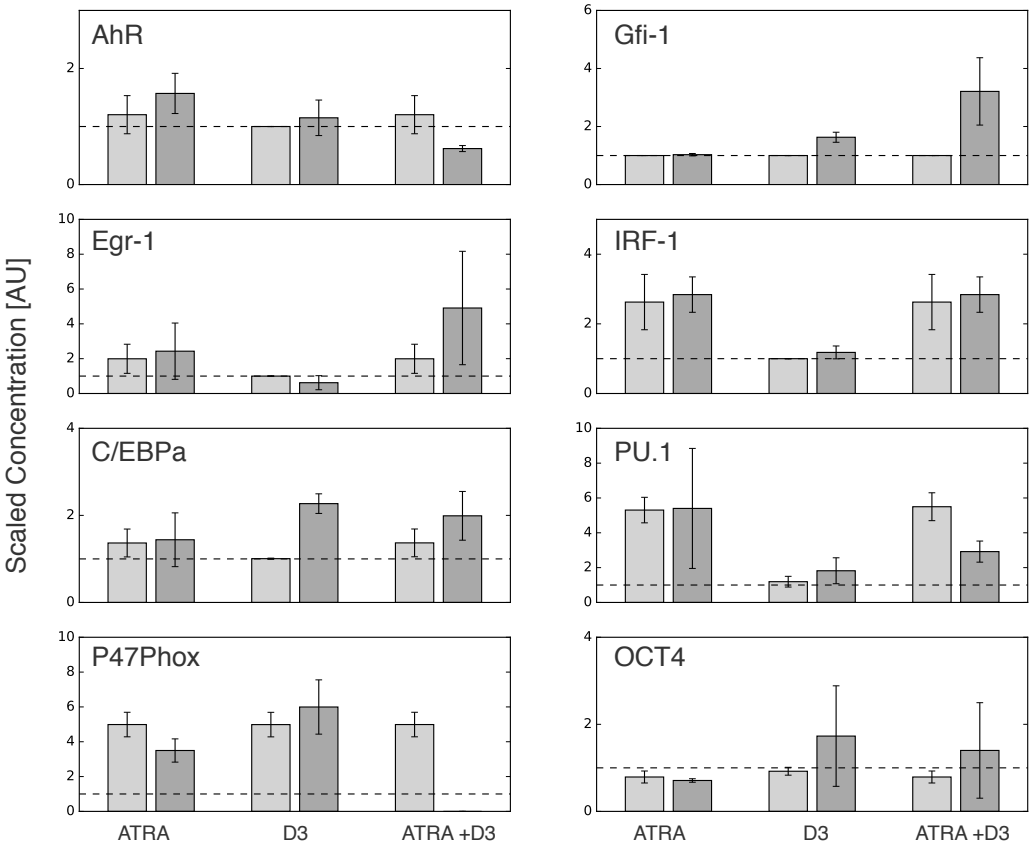

Figure 3.2: ATRA, D3, and combination induced expression of key proteins in the gene regulatory network at 48 hours after induction. The error bars represent standard errors, 491 samples for simulated data, and 3 samples for experimental measured data. The dotted horizontal line indicated the baseline level of protein at time 0hr.

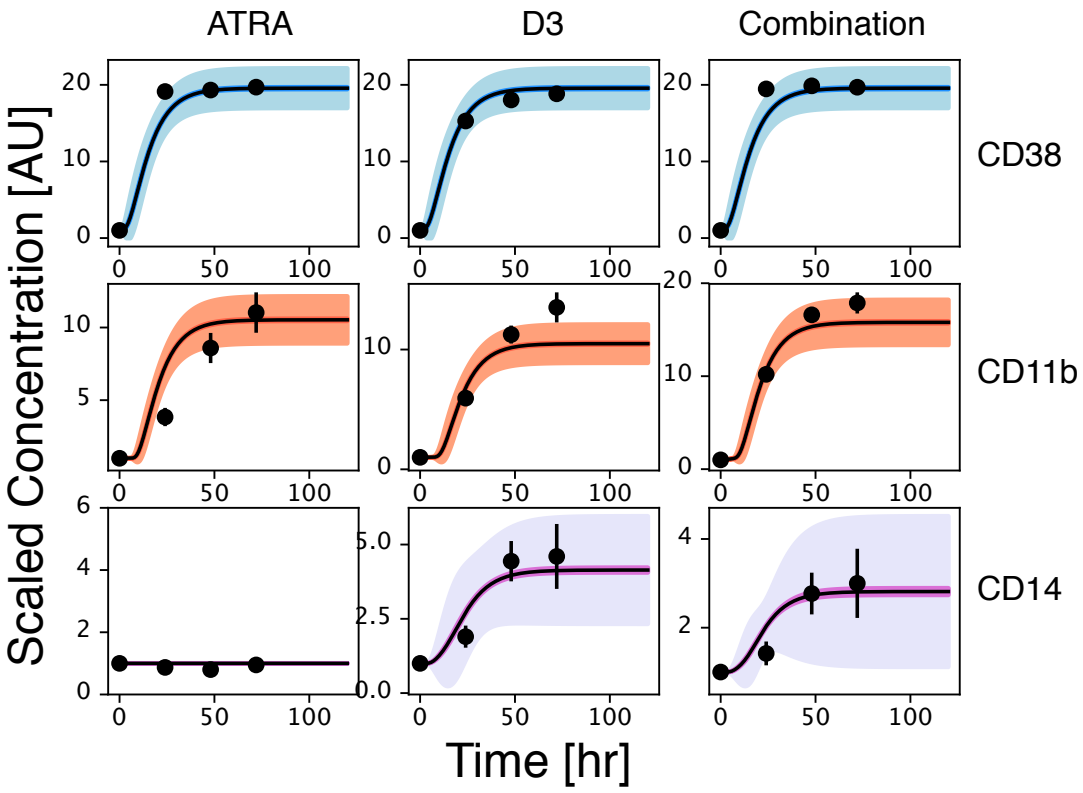

Figure 3.3: Dynamic phenotypic markers of ATRA, D3, and combination induced differentiation of HL60. The simulation captured time-course experimental measurements of CD38 (blue), CD11b (red), and CD14 (purple). The points are experimental measurements with error bars representing standard error. The black lines are the simulated average of the 491 parameter set ensemble, with the darker colored band representing 99 % standard error of the simulated mean, and the lighter colored bands meaning 99 % confidence interval.

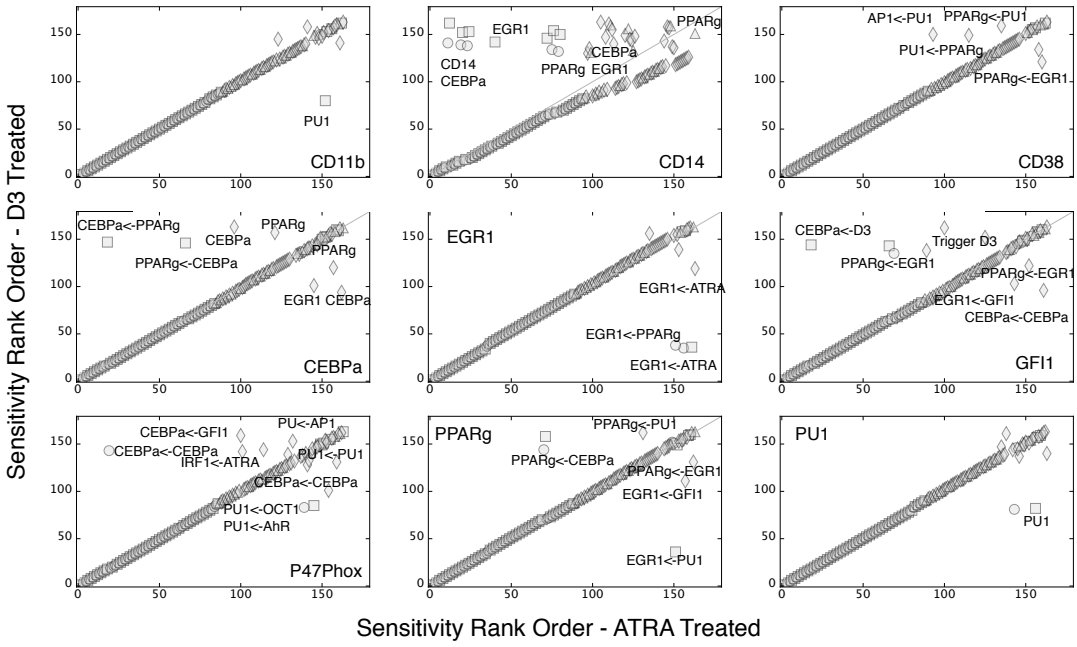

Figure 3.4: Sensitivity analysis comparing ATRA and D3 induced HL60 differentiation. The parameters are ranked ordered from low to high sensitivity with the parameters that are more sensitive for ATRA and D3 treatment fall on the right and left side of the 45 degree line respectively. The diamonds represent the weight parameters (W), the squares are saturation constants (K_s), and the circles are cooperativity coefficients (n).

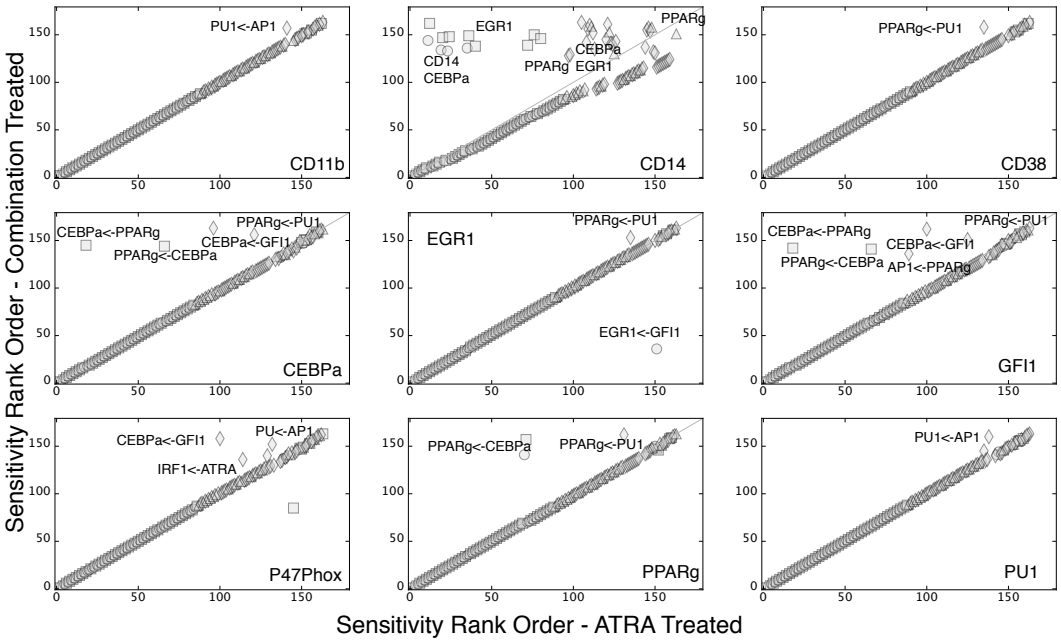

Figure 3.5: Sensitivity analysis comparing ATRA and D3 induced HL60 differentiation. The parameters are ranked ordered from low to high sensitivity with the parameters that are more sensitive for ATRA and D3 treatment fall on the right and left side of the 45 degree line respectively. The diamonds represent the weight parameters (W), the squares are saturation constants (K_s), and the circles are cooperativity coefficients (n).

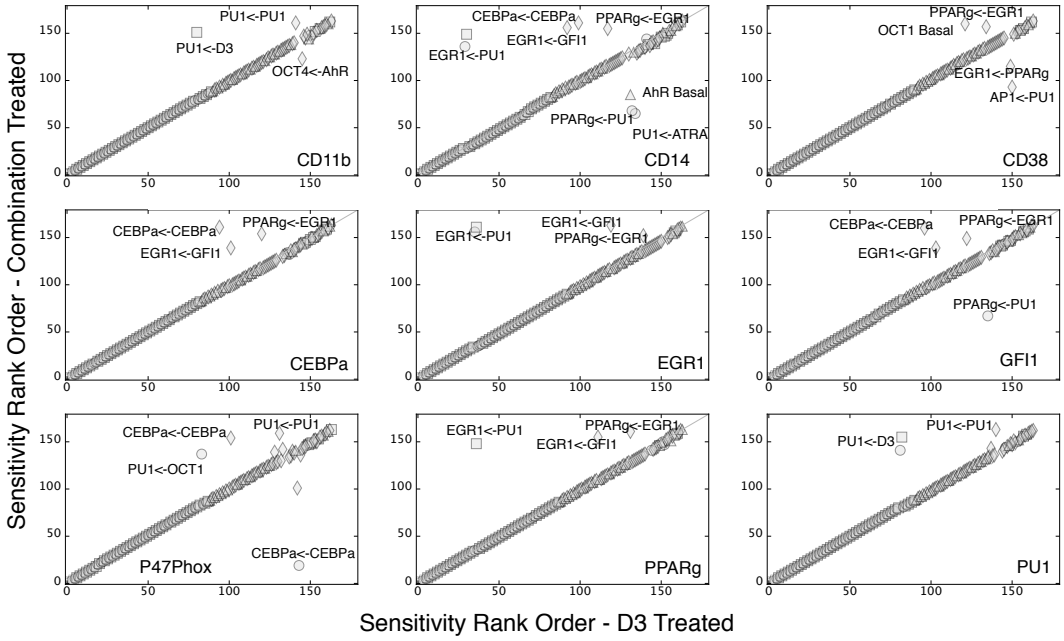

Figure 3.6: Sensitivity analysis comparing D3 and combination induced HL60 differentiation. The parameters are ranked ordered from low to high sensitivity with the parameters that are more sensitive for D3 and combination treatment fall on the right and left side of the 45 degree line respectively. The diamonds represent the weight parameters (W), the squares are saturation constants (K_s), and the circles are cooperativity coefficients (n).

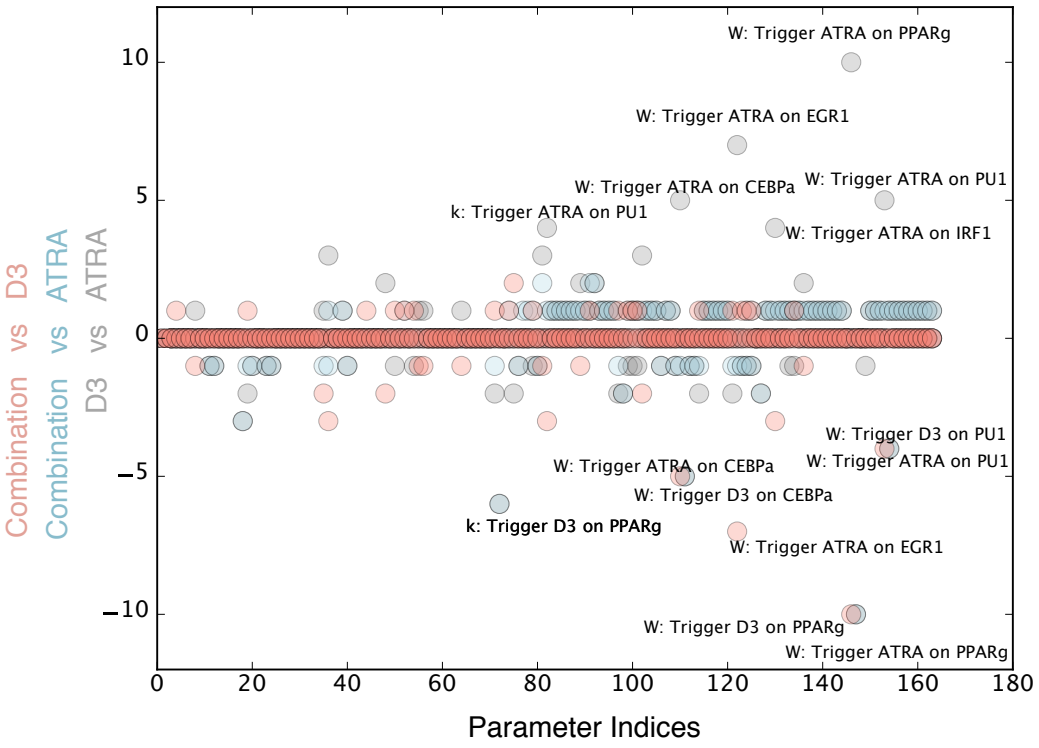

Figure 3.7: The frequency of off-axis appearance of ranked-order sensitivity comparison between different treatments. The grey circle indicate ATRA vs D3 comparison, with ATRA sensitive parameters in the positive region and D3 sensitive parameters in the negative. The blue data is the ATRA vs combination comparison, while the red is the D3 vs combination comparison.

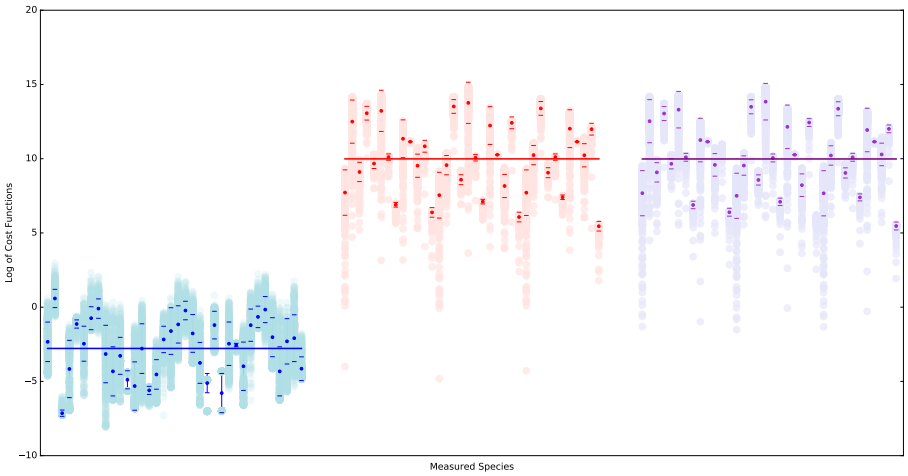

Figure 3.8: Log of the 36 cost functions associated all three treatment plotted for three different 495-parameter-set-ensemble. Blue is the ensemble generated using parameter estimation, red and purple are ensembles generated randomly and using Sobol sequences respectively.

Table 3.1: Myelomonocytic transcription factor connectivity used in the signal integration and phenotype modules.

Effector	Effect	Target	Source
RARα	+	RARα	[192]
	+	PU.1	[175]
	+	C/EBPα	[76]
	+	IRF-1	[160]
	-	Oct4	[220]
	+	CD38	[53]
	+	p21	[155]
	+	AhR	[23]
	+	Egr-1	[7]
PPARγ	+	C/EBPα	[195]
	+	IRF-1	[232]
	+	Oct1	[20]
	-	AP-1	[47]
	-	E2F	[4]
	-	Egr-1	[66]
	+	CD38	[211]
	+	CD14	[221]
	+	p21	[86]
	-	p47Phox	[236]
PU.1	-	PPARγ	[49]
	+	PU.1	[28]
	+	AP-1	[215]
	+	Egr-1	[142]
	+	CD11b	[185]
	+	p21	[257]
	+	p47Phox	[149]
C/EBPα	+	PPARγ	[195]

	+	PU.1	[43]
	+	C/EBPα	[227]
	+	Gfi-1	[153]
	-	E2F	[44]
	+	CD14	[187]
	+	p21	[89]
IRF-1	+	CD38	[9]
	+	p21	[189]
	-	PU.1	[42]
	-	C/EBPα	[54]
	-	E2F	[54]
	-	Egr-1	[142]
	-	p21	[54]
Oct1	+	PU.1	[29]
AP-1	-	PPARγ	[47]
	+	PU.1	[11]
	+	p21	[125]
E2F	+	E2F	[123]
Egr-1	+	PPARγ	[77]
	-	Gfi-1	[163]
	+	CD14	[27]
AhR	+	AP-1	[217]
	+	IRF-1	[206]
	-	Oct4	[22]
	-	PU.1	[14]

CHAPTER 4

SRC FAMILY KINASE INHIBITOR BOSUTINIB ENHANCES RETINOIC ACID-INDUCED DIFFERENTIATION OF HL-60 LEUKEMIA CELLS

4.1 Abstract[1]

A subtype of acute myeloid leukemia (AML), acute promyelocytic leukemia, has been treated with all-trans retinoic acid (RA) for several decades; the drug induces differentiation of cancerous cells. However, many patients relapse and cancers become RA resistant. Combination therapy is a way to maintain and potentially broaden efficacy of the treatment. Using the RA-responsive AML cell line HL-60, we tested the efficacy of the Src family kinase (SFK) inhibitor bosutinib (B) on RA-induced differentiation. We found that co-treatment with RA and B enhanced differentiation evidenced by increased CD11b expression, G1/G0 cell cycle arrest, and respiratory burst. Expression of the SFK members Fgr and Lyn was enhanced, while SFK activation was inhibited. Phosphorylation of several sites of c-Raf was increased, and expression of AhR and p85 PI3K was enhanced. Expression of c-Cbl and mTOR was decreased. Our study suggests that SFK inhibition enhances RA-induced differentiation and may have therapeutic value.

[1]This work is published and adopted from Leukemia Lymphoma DOI:10.1080/10428194.2018.1452213

4.2 Introduction

Acute myeloid leukemia (AML), the most common form of acute leukemia in adults, is characterized by defective differentiation and excessive accumulation of proliferatively active progenitor cells in bone marrow and blood [105, 111]. One of the most successful treatments of AML is of its subtype, acute promyelocytic leukemia (APL). APL is characterized by a t(15,17) cytogenetic marker seminal to a block of leukocyte differentiation at the promyelocyte stage and accounts for approximately 10–15% of all AML cases [107, 246]. Once considered one of the most lethal forms of acute leukemia, the advent of all trans-retinoic acid (RA) and arsenic trioxide therapy has revolutionized treatment of this disease [246]. Now, APL is characterized by complete remission rates of 90% and cure rates of around 80% [38]. While RA is effective in treating APL, it is not effective in treating other subtypes of AML. Moreover, many APL patients who initially respond to RA treatment relapse; subsequent RA treatment is ineffective due to RA-resistance [245, 74]. It is therefore of great interest to combine RA with other agents, such as other differentiation-inducing compounds or kinase inhibitors, in order to promote RA-induced differentiation of APL as well as non-APL AML [116]. Src family kinases (SFKs) are a group of enzymes that are important in leukemia cell proliferation, survival, adhesion, and differentiation [111]. SFKs have been shown to positively regulate MAPK signaling, cell proliferation, and contribute to cell transformation [135]. SFKs are overexpressed in many cancer types and are typically associated with acute and chronic myeloid malignancies and metastasis [156, 136]. Lyn has been found to be the primary active SFK expressed in AML cells [52, 141]. SFK inhibition has been effective in slowing leukemic cell growth [**?**]. It is therefore of interest to determine the

impact of SFK expression and activity on RA-induced differentiation therapy. SFK inhibitors PP2 and dasatinib have been reported to enhance RA-induced differentiation [141, 173, 34, 128, 118]. However, some reports demonstrate that SFKs positively regulate RA-induced differentiation. It is ergo unclear what role SFKs have in regulating RA-induced differentiation. One focus of interest is the impact of SFK inhibitors on the MAPK pathway, where different signaling/phosphorylation signatures on the same targeted signaling molecules can cause different outcomes. Defining these signatures is important to understanding the differential signaling attributes that might be exploited for therapeutic intervention. RA has been found to elicit MAPK pathway activation necessary for HL-60 cell differentiation and growth arrest [254]. PP2 and dasatinib co-treatments with RA were found to affect the Raf/MEK/ERK axis by upregulating c-Raf pS259 while not impacting MEK or ERK expression or phosphorylation [34]. The results are somewhat counter-intuitive, but suggest signaling events of importance to inducing differentiation of the leukemic cells, and they motivate targeting SFKs to probe for signaling attributes driving differentiation. Bosutinib (B) is a second-generation SFK inhibitor that has been used clinically to treat phases of intolerant or resistant Philadelphia chromosome (t(19,21)/Bcr-Abl) – positive chronic myeloid leukemias (CML) that do not harbor the T315I or V299LABL kinase domain mutations [2]. While B treatments have been studied extensively in CML, its effects on AML have not been characterized. B's clinical usage to treat CML and benign toxicity profile make it an attractive candidate for combination treatment on AML [2, 39]. We examined the effects of RA/B co-treatments on WT HL-60 cells to determine the effects of bosutinib on HL-60 differentiation and the MAPK and mTOR pathways. We found that RA/B enhanced RA-induced expression of the differentiation marker CD11b at

48 and 72 h and G0/G1 arrest at 48 h. Induced oxidative metabolism and associated marker p47phox, a component of the NADPH oxidative metabolism machine, were also enhanced with RA/B compared to RA. With RA/B treatment, SFK members Fgr and Lyn saw an increase in expression, while pan-SFK activation decreased. B also enhanced RA-induced c-Raf phosphorylation at S259, S621, and the C-terminal domain. B decreased expression of c-Cbl and mTOR compared to RA alone. B enhanced RA-induced expression of AhR and p85 PI3K. Hence B affected several known key signaling molecule regulators of RA-induced leukemic cell differentiation. Like dasatinib, combined RA/B therapy may be useful in future AML treatments.

4.3 Materials and Methods

Cell culture and treatments Reagents, unless specified otherwise, were purchased from commercial suppliers in the highest purity available. HL-60 human myeloblastic leukemia cells, derived from the original patient isolates were maintained in this laboratory in RPMI 1640 supplemented with 5% fetal bovine serum (GE Healthcare, Chicago, IL) and 1x antibiotic/antimycotic (Thermo Fisher Scientific, Waltham, MA) in a 5% CO2 humidified atmosphere at 37C. The cells used were certified as mycoplasma free HL-60 by Bio-Synthesis, Lewisville, TX, in August 2017. Viability was monitored by 0.2% trypan blue (Invitrogen, Carlsbad, CA) exclusion and routinely exceeded 95%. Experimental cultures were initiated at a density of 0.1106 cells/ml. There were three treatment regimens studied: (1) untreated, (2) RA, and (3) RA/B. All-trans retinoic acid (RA) (Sigma, St. Louis, MO) was added from a 5 mM stock solution in 100% ethanol to a final concentration of 1 in culture. Bosutinib (Sigma, St. Louis, MO) was

used from a stock of 5mM in DMSO to make the final concentrations in culture indicated. Flow cytometric phenotypic analysis Immunostaining for CD11b and CD38 was performed as previously described and analyzed using a Becton Dickinson LSR II flow cytometer (San Jose, CA). Gating was set to exclude 95% of the untreated wild-type HL-60 samples. Propidium iodide (PI) cell cycle analysis was performed as previously described. Respiratory burst quantification Respiratory burst quantification was performed as previously described and analyzed using a Becton Dickinson LSR II flow cytometer (San Jose, CA). Gating was set to exclude 95% of the samples not induced with TPA. The shift in fluorescence intensity in response to TPA was used to determine the percent cells with the capability to generate inducible oxidative metabolites. Antibodies Lyn, Fgr, pY416-SFK, AhR, p47phox, mTOR, pS259c-Raf, pS621c-Raf, GAPDH, horseradish peroxidase anti-mouse and anti-rabbit antibodies were from Cell Signaling (Danvers, MA, USA). c-Cbl (C-15) and AhR (H211) were from Santa Cruz Biotechnology (Santa Cruz, CA). Western blot analysis Cells were pelleted, washed twice with PBS, and lysed with ice cold M-PER Mammalian Protein Extraction Reagent (Pierce, Rockford, IL) with protease and phosphatase inhibitor cocktails (Sigma, St. Louis, MO). Samples were incubated overnight at 80C and debris was pelleted. Protein concentration was determined using the Pierce BCA Protein Assay according to the manufacturer's protocol. Lysate was subjected to standard SDS-PAGE, using 25of lysate per lane under denaturing conditions. Membranes were blocked with 5% dry nonfat milk in PBS and probed with antibodies described above. Enhanced chemiluminescence reagent (GE Healthcare, Pittsburg, PA) was used for detection. Films were scanned and quantitated using NIH Image J. Statistical analysis Statistics were analyzed using Microsoft Excel. Means of treatment groups of interest were compared using

paired sample T-tests. The data represent the means of three repeats S.E.M. A p-value of ¡0.05 was considered significant.

4.4 Results

RA/B enhances CD11b and G1/G0 arrest We first determined the effect of B on RA-induced differentiation in HL-60 cells by comparing differentiation markers of cells treated with RA alone or in combination with B (RA/B) over a 72 h treatment period: we assessed CD38 and CD11b expression, G1/G0 cell cycle arrest, and population growth. We chose to use 0.25 B based on an initial dose-response experiment, included in the supplementary information (Figure S1). A previous study also found that this dosage of B inhibited cancer cell migration and invasion [237]. We measured expression of two cell surface markers, CD38 and CD11b, using flow cytometry. CD38 expression was nearly 100% in both RA and RA/B treated cells at 24, 48, and 72 hours (Fig. 1A). CD38 mean expression per cell also yielded no difference between RA and RA/B at the three time points (Fig. 1B). We measured CD11b, a later differentiation marker of RA-induced differentiation, at 48 h and 72 h. RA/B significantly increases CD11b expression at 72 h (Fig. 1C, p ¡ 0.05, n=3). Mean expression per cell of CD11b also shows a significant increase with RA/B treatment at both 48 h and 72 h (Fig. 1D, p ¡ 0.05, n=3). At 48 h but not 72 h, RA/B showed significant enhancement of G0/G1 arrest compared to RA alone (Fig. 1E, p ¡ 0.05, n=4). Addition of B to the cells did not result in any apparent toxicity compared to RA, as both RA/B and RA treatments yield similar cell growth patterns over the 72 h treatment period (Fig. 1F). Both RA and RA/B induce growth retardation compared to untreated cells (Fig. 1F). RA/B increases RA-induced respiratory burst activity

To determine the effects of RA/B on a functional differentiation marker, respiratory burst, we measured inducible reactive oxygen species production (ROS) by flow cytometry at 72 h. RA caused an increase in ROS compared to untreated cells, and RA/B treated cells show a significant further increase in oxidative metabolism compared to RA alone (Fig. 2A, p ¡ 0.05, n=3). To corroborate the ROS assay, we also measured expression of a subunit of the NADPH oxidase complex that produces the respiratory burst, p47phox (21,22). After a 48 h treatment period, we collected cell lysate and performed western blots for p47phox (Fig. 2B). Expression was greater in both RA and RA/B compared to untreated cells, and RA/B further enhanced p47phox levels compared to RA alone (Fig. 2B). RA/B increases RA-induced SFK expression, but decreases SFK activation As B is an SFK inhibitor and the SFK members Fgr and Lyn have been shown to be upregulated by RA treatment in HL-60 cells [107, 34, **?**], we assessed levels and phosphorylation of these members by western blot. We collected lysate following a 48 h treatment period. Both RA and RA/B increased Fgr and Lyn levels compared to untreated cells, and RA/B induced further increases of Fgr and Lyn levels compared to RA alone (Fig. 3A, 3B). We measured SFK activation using a pan-SFK antibody that detects active site (Y416) phosphorylation in all family members, including Fgr and Lyn. While treatment with RA and RA/B both caused in increase in SFK phosphorylation compared to untreated cells, treatment with RA/B decreased SFK phosphorylation compared to RA alone (Fig. 3, A-C). RA/B augments RA-induced c-Raf phosphorylation We also examined phosphorylation of several sites of c-Raf previously shown to drive RA-induced differentiation. [256]. Since SFKs can regulate MAPK signaling, we analyzed the effect of B on RA-induced c-Raf activation. We evaluated the phosphorylation status of c-Raf regulatory phosphorylations at S259, S621, and the

c-Raf C-terminal domain by western blot following a 48 h treatment period. RA increases the amount of c-Raf pS259, c-Raf pS621, and c-Raf C-terminal domain phosphorylation. We observed modest enhancements of levels of each of these in cells treated with RA/B compared to RA alone (Fig. 4, A-C). RA/B inhibits c-Cbl and mTOR expression and enhances AhR and p85 PI3K expression We next assessed levels of several proteins associated with RA-induced differentiation, AhR, c-Cbl, and p85 PI3K, as well as one known to be aberrantly activated in AML, mTOR [119, 52, **?**, 35]. After a 48 h treatment period, we collected cell lysate and analyzed expression of these proteins via western blot. C-Cbl expression was upregulated in cells treated with RA and RA/B, but RA/B inhibited c-Cbl expression compared to cells treated with RA alone (Fig. 5A). mTOR expression followed a similar pattern (Fig. 5B). AhR and p85 PI3K levels, however, were further upregulated in RA/B-treated cells compared to RA-treated cells (Fig 5C, 5D).

4.5 Discussion

We sought to provide an initial characterization of the effects of treatment with B on RA-induced differentiation of HL-60 cells. We observed that RA/B treatment enhanced several myeloid lineage differentiation markers compared with RA treatment alone: CD11b expression, G1/G0 cell cycle arrest, and respiratory burst, a functional marker of mature myeloid series cells [115]. Expression of CD38, another feature of RA-induced differentiation, was not affected by the addition of B. Recent reports from our laboratory, however, suggest that CD38 may not play an important role in the differentiation process (Sci Rep paper, Oncotarget paper). Since addition of B enhanced phenotypic and func-

tional markers of RA-induced differentiation, we explored its effects on activation of the key signaling protein c-Raf. Cells treated with RA/B displayed increased phorphorylated c-Raf levels. The particular phosphorylation sites assayed, S259, S621, and the C-terminal domain, are all associated with active c-Raf in HL-60 cells and thus drive RA-induced differentiation [256]. Our western blotting results are consistent with our phenotypic results and previous findings that RA co-treatments with PP2 or dasatinib also increased c-Raf phosphorylation, namely c-Raf pS259 and c-Raf pS621, in HL-60 cells [34]. These findings are congruous with the notion that c-Raf is a downstream target of SFK inhibitors like B. Given that B is a SFK inhibitor, we assessed its effects on two SFK members, Fgr and Lyn, the primary active SFK expressed in AML cells [52, 141]). As anticipated, we saw that RA/B treatment decreased levels of phosphorylated Y416, a mark of activated SFKs. We did, however, observe increased levels of Fgr and Lyn in RA/B compared to RA alone; increase in expression could perhaps be an attempt to compensate for the inhibited activity of the proteins. These results are consistent with previous findings with RA/dasatinib co-treatments [34]. Dasatinib, like B, is a second generation SFK inhibitor that is used clinically to treat chronic myeloid leukemia (CML). B has a lower toxicity profile compared to dasatinib and is used in patients with CML who have shown intolerance to dasatinib [124]. c-Cbl interacts with CD38 to promote RA-induced differentiation and G1/G0 arrest of HL-60 cells [**?**]. As previously reported, c-Cbl expression is upregulated by RA treatment; RA/B treatment partially inhibits RA-induced upregulation. Since c-Cbl expression has been shown to drive RA-induced differentiation, this is surprising, but it is possible that B drives RA-induced differentiation via other pathways such as c-Raf activation. The mTOR/p70S6K/4EBP1 pathway, a driver of cellular

anabolism/metabolism, is constitutively activated in AML cells, leading to increased cell proliferation and inhibition of differentiation [52]. mTOR is thought to be downstream of the pathways implicated above to regulate differentiation. Dasatinib inactivates the mTOR pathway, correlating with an improvement in cell survival in the AML-derived cell line PVTL-1 [177]. Lyn is hypothesized to serve as the intermediary; Lyn stimulates mTOR expression, and is inhibited by dasatinib [177]. Lending credence to this, silencing Lyn was found to inhibit the mTOR pathway in primary AML cells, and PP2 mimics rapamycin, a selective inhibitor of mTOR [52]. To our knowledge, the effects of SFK inhibitors on the mTOR pathway in APL are unknown. It was hence of interest to determine if B has an inhibitory effect on mTOR in an RA-responsive AML cell line as found in other forms of AML. Both RA and RA/B induced greater mTOR expression compared to untreated cells, however, RA/B decreased mTOR expression compared to RA (Fig. 5B). It appears that a decrease in the activated SFKs yields an inhibitory effect on the mTOR pathway in RA-treated cells. This inhibitory effect on the mTOR pathway suggests part of the process by which B enhances differentiation compared to RA alone. We note that MAPK signaling pathway activation is promoting differentiation and cell cycle arrest in RA-treated HL-60 cells, yet it is also thought to drive mTOR and, consequentially, cellular anabolism to promote proliferation. Hence, RA appears to be redirecting these pathways to support differentiation instead of proliferation. The mechanism of this is a profound but yet unresolved question in leukemic cell differentiation. AhR is known to drive differentiation and to be linked to phosphorylated c-Raf in an RA-activated signalsome (8). We found that RA treatment increased AhR levels, consistent with previous studies (Fig. 5C) [107, 190, 177]. Co-treatment with B further increased AhR expression, consistent with the increase in phos-

phorylated c-Raf expression in RA/B treated samples (Fig. 3, Fig. 5C). p85 PI3K activity is upregulated during granulocytic maturation [35]. It was therefore of interest to determine the effect of B on p85 PI3K expression; the protein exhibited an expression pattern similar to that of AhR, as B caused an increase in RA-induced upregulation of p85 PI3K expression. Interestingly, while both CD38 and c-Cbl are known to interact with p85 PI3K [35], RA/B treatment did not affect the RA-induced expression of these proteins in the same fashion. B, like other SFK inhibitors, has promising effects on enhancing RA-induced differentiation in HL-60 cells. These effects include modulation of the SFK members Fgr and Lyn and impacts on the MAPK and mTOR pathways. Further studies may elucidate the effects of RA/B co-treatments on RA-resistant APL cell lines or other myeloid leukemias. B is already in clinical trials as a cytotoxic chemotherapeutic agent, but our results suggest it is potentially useful as an agent used in combination therapy with RA for differentiation therapy of AML.

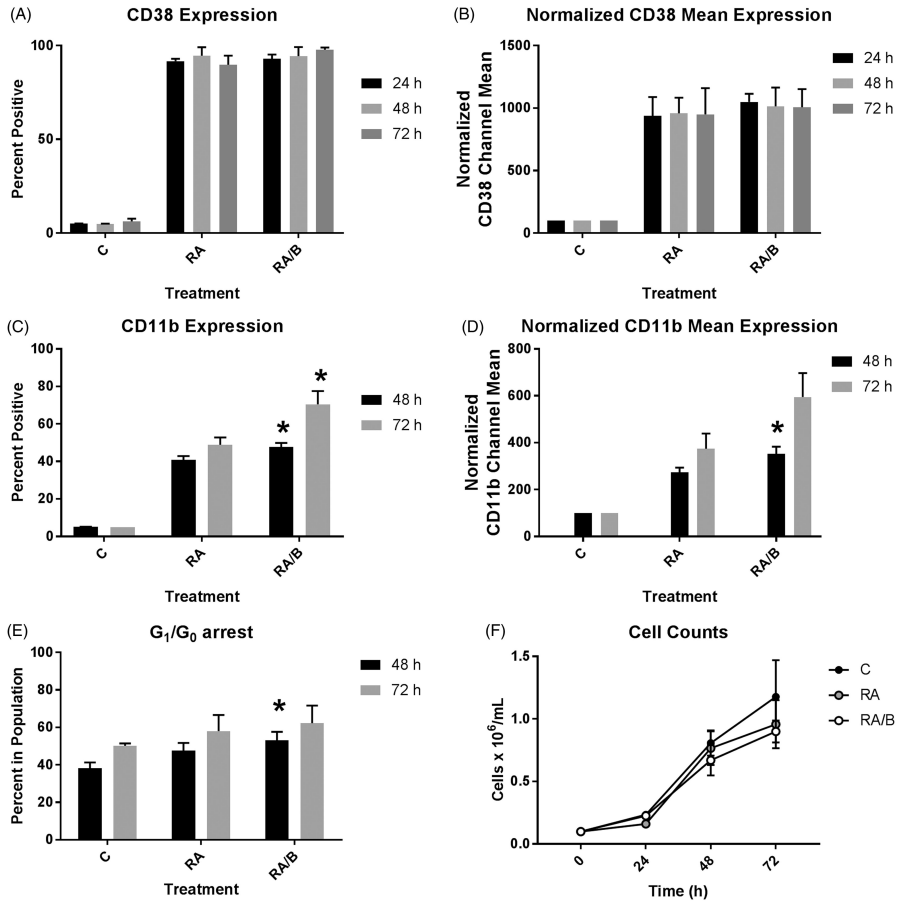

Figure 4.1: Phenotypic analysis of HL-60 cells treated with RA/bosutinib. (A) HL-60 cells were cultured in the presence of 1RA or 1RA and 0.25bosutinib as indicated. CD38 expression was assessed by flow cytometry following 24, 48, and 72h treatment periods. Gating to discriminate positive cells was set to exclude 95% of untreated controls (n=3). Error bars indicate SEM. (B) Normalized means of CD38 expression per cell at 24, 48, or 72h (n=3). (C) CD11b expression was assessed by flow cytometry at 48 and 72h (n=3). *p¡.05 comparing RA-treated samples to RA/bosutinib-treated samples. Two-tailed paired-sample t-tests were used to determine significance. (D) Normalized means of CD11b expression per cell at 48and 72h (n=3). *p¡.05 comparing RA-treated samples to RA/bosutinib-treated samples. (E) Cell cycle distribution showing the percentage of cells in G1/G0 was analyzed using flow cytometry with propidium iodide staining at 24, 48, and 72h (n=4). *p¡.05 comparing RA-treated samples to RA/bosutinib-treated samples. (F) Cell counts were taken at 24, 48, and 72h using a hemocytometer and 0.2% Trypan Blue exclusion staining (n=3).

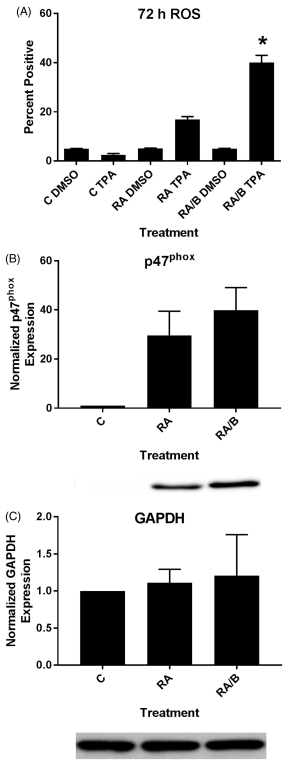

Figure 4.2: HL-60 cells treated with RA/bosutinib displayed enhanced respiratory burst and p47phox expression. (A) HL-60 cells were cultured in the presence of 1RA or 1RA and 0.25bosutinib as indicated. Respiratory burst was analyzed by measuring inducible reactive oxygen species (ROS) production by flow cytometry using the 2,7-dichlorofluorescein (DCF) assay. Gates to determine percent increase of expression with treatment were set to exclude 95% of the DMSO-treated control population for each culture condition; TPA-treated samples show induced ROS (n=3). Error bars indicate SEM. *p¡.05 comparing RA-treated samples to RA/bosutinib-treated samples. Two-tailed paired-sample t-tests were used to determine significance. (B) HL-60 cells were cultured for 48h in the presence of 1RA or 1RA and 0.25B as indicated and whole cell lysate was collected. Twenty five microgram of lysate per lane was run. Western blots of PAGE-resolved lysates were probed for p47phox (n=3). Films were scanned and bands of interest were quantified using NIH ImageJ. Error bars indicate SEM. A representative blot, cropped to show only the band of interest, is included. (C) Western blots of GAPDH were used as loading controls following the procedure described above.

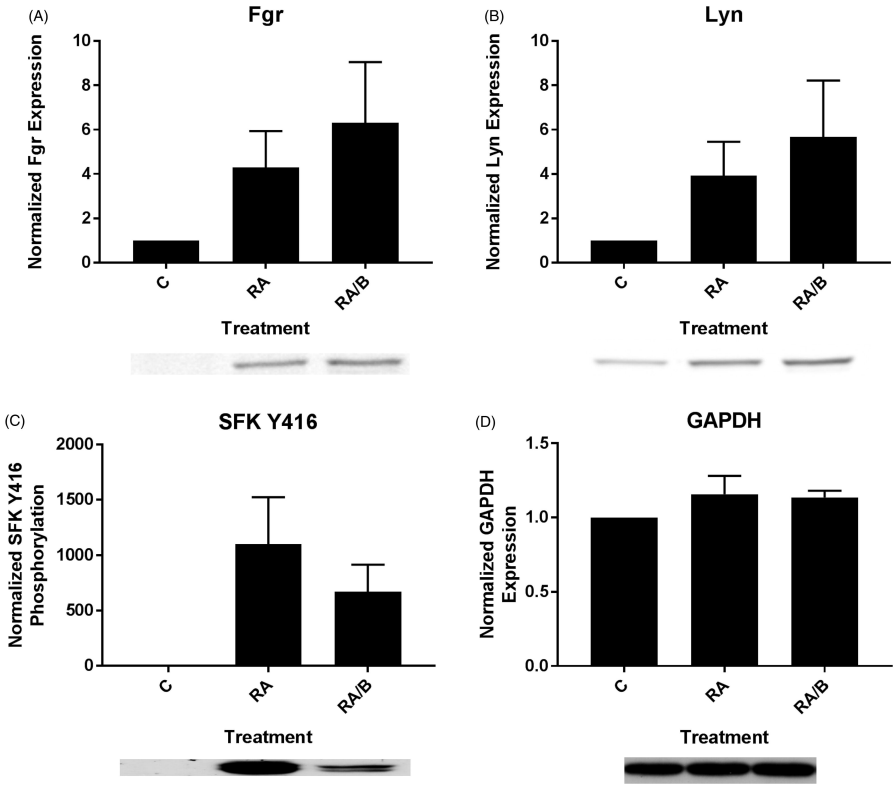

Figure 4.3: Bosutinib enhances RA-induced SFK expression and diminishes SFK phosphorylation. (A) HL-60 cells were cultured for 48h in the presence of 1RA or 1RA and 0.25bosutinib (B) as indicated and whole cell lysate was collected. Twenty five microgram of lysate per lane was run. Western blots of PAGE-resolved lysates were probed for Fgr (n=3). Films were scanned and bands of interest were quantified using NIH ImageJ. Error bars indicate SEM. A representative blot, cropped to show only the band of interest, is included. (B) Western blots of Lyn following the procedure described above. (C) Western blots of phosphorylated pan-Y416 SFK following the procedure described above. (D) Western blots of GAPDH were used as loading controls following the procedure described above.

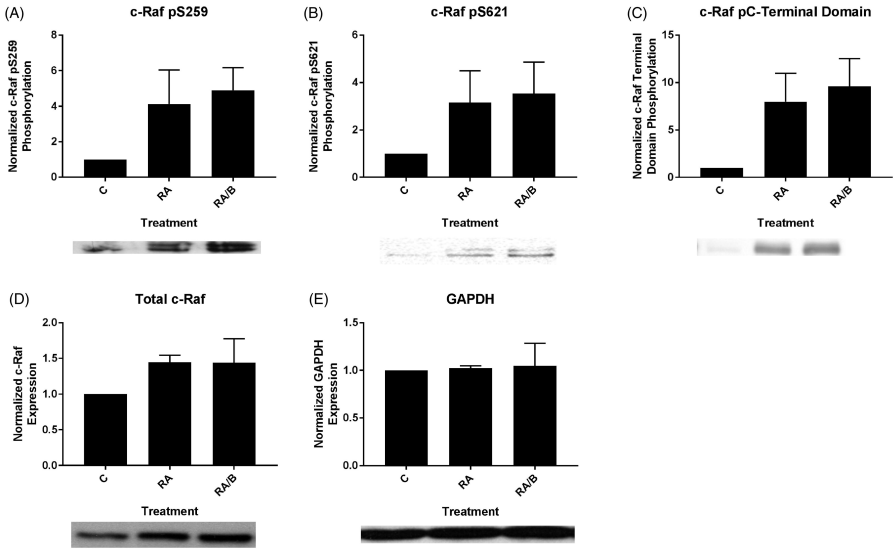

Figure 4.4: Bosutinib enhances levels of RA-induced phosphorylated c-Raf. (A) HL-60 cells were cultured for 48h in the presence of 1RA or 1RA and 0.25bosutinib as indicated and whole cell lysate was collected. Twenty five microgram of lysate per lane was run. Western blots of PAGE-resolved lysates were probed for c-Raf pS259 (n=3). Films were scanned and bands of interest were quantified using NIH ImageJ. Error bars indicate SEM. A representative blot, cropped to show only the band of interest, is included. (B) Western blots of c-Raf pS621 following the procedure described above. (C) Western blots of c-Raf pC-terminal domain following the procedure described above. (D) Western blots of total c-Raf following the procedure described above. (E) Western blots of GAPDH were used as loading controls following the procedure described above.

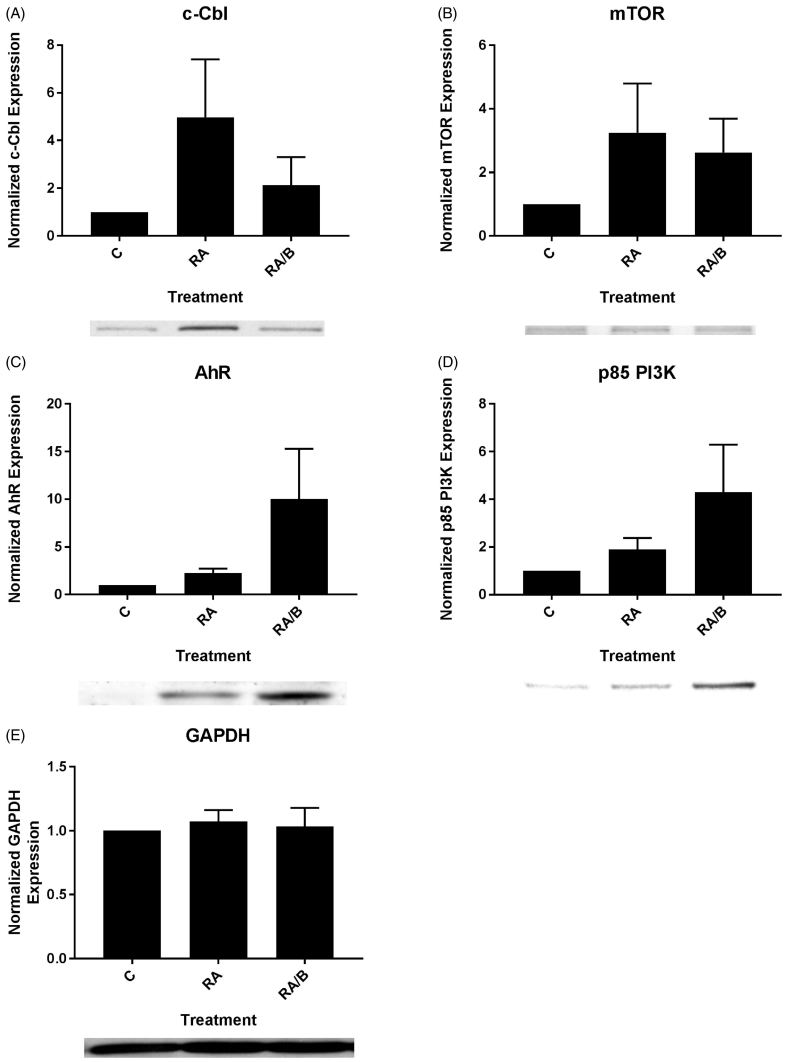

Figure 4.5: The effect of bosutinib on c-Cbl, mTOR, AhR, and p85 PI3K. (A) HL-60 cells were cultured for 48h in the presence of 1RA or 1RA and 0.25bosutinib (B) as indicated and whole cell lysate was collected. Twenty five microgram of lysate per lane was run. Western blots of PAGE-resolved lysates were probed for c-Cbl (n=3). Films were scanned and bands of interest were quantified using NIH ImageJ. Error bars indicate SEM. A representative blot, cropped to show only the band of interest, is included. (B) Western blots of mTOR following the procedure described above. (C) Western blots of AhR following the procedure described above. (D) Western blots of p85 PI3K following the procedure described above. (E) Western blots of GAPDH were used as loading controls following the procedure described above.

CHAPTER 5

DYNAMIC SEQUENCE SPECIFIC CONSTRAINT-BASED MODELING OF CELL-FREE PROTEIN SYNTHESIS

5.1 Abstract[1]

Cell-free protein expression has emerged as an important approach in systems and synthetic biology, and a promising technology for personalized point of care medicine. Cell-free systems derived from crude whole cell extracts have shown remarkable utility as a protein synthesis technology. However, if cell-free platforms for on-demand biomanufacturing are to become a reality, the performance limits of these systems must be defined and optimized. Toward this goal, we modeled *E. coli* cell-free protein expression using a sequence specific dynamic constraint based approach in which metabolite measurements were directly incorporated into the flux estimation problem. A cell free metabolic network was constructed by removing growth associated reactions from the *i*AF1260 reconstruction of K-12 MG1655 *E. coli*. Sequence specific descriptions of transcription and translation processes were then added to this metabolic network to describe protein production. A linear programming problem was then solved over short time intervals to estimate metabolic fluxes through the augmented cell free network, subject to material balances, time rate of change and metabolite measurement constraints. The approach captured the biphasic cell free production of a model protein, chloramphenicol acetyltransferase. Flux variability analysis suggested that cell free metabolism was potentially robust; for example, the rate of protein production could be met by flux through the

[1]This work is published and adopted from Processes https://doi.org/10.3390/pr6080132

glycolytic, pentose phosphate, or the Entner-Doudoroff pathways. Variation of the metabolite constraints revealed central carbon metabolites, specifically upper glycolysis, TCA cycle, and pentose phosphate, to be the most effective at training a predictive model, while energy and amino acid measurements were less effective. Irrespective of the measurement set, the metabolic fluxes (for the most part) remained unidentifiable. These findings suggested dynamic constraint based modeling could aid in the design of cell free protein expression experiments for metabolite prediction, but the flux estimation problem remains challenging. Further, while we modeled the cell-free production of only a single protein in this study, the sequence specific dynamic constraint based modeling approach presented here could be extended to multi-protein synthetic circuits, RNA circuits or even small molecule production.

5.2 Introduction

Cell-free protein expression has become a widely used research tool in systems and synthetic biology, and a promising technology for personalized point of use biotechnology [188]. Cell-free systems offer many advantages for the study, manipulation and modeling of metabolism compared to *in vivo* processes. Central amongst these is direct access to metabolites and the biosynthetic machinery, without the interference of a cell wall or the complications associated with cell growth. This allows us to interrogate (and potentially manipulate) the chemical microenvironment while the biosynthetic machinery is operating, potentially at a fine time resolution. Cell-free protein synthesis (CFPS) systems are arguably the most prominent examples of cell-free systems used today [121]. However, CFPS in crude *E. coli* extracts has been used previously to explore fundamen-

tal biological questions. For example, Matthaei and Nirenberg used *E. coli* cell free extracts to decipher the genetic code [167, 179]. Later, Spirin and coworkers continuously exchanged reactants and products in a CFPS reaction, which improved protein production. However, while these extracts could run for up to tens of hours, they could only synthesize a single product and were likely energy limited [213]. More recently, energy and cofactor regeneration in CFPS has been significantly improved; for example ATP can be regenerated using substrate level phosphorylation [134] or even oxidative phosphorylation [121]. Today, cell-free systems are used in a variety of applications ranging from therapeutic protein production [157] to synthetic biology [99, 188]. There are also several CFPS technology platforms, such as the PANOx-SP and Cytomin platforms developed by Swartz and coworkers [122, 121], and the TX/TL platform of Noireaux [79]. Taken together, CFPS is a promising technology for protein production. However, if CFPS is to become a mainstream technology for applications such as point of care biomanufacturing, we must first understand the performance limits of these systems, and eventually optimize their yield and productivity. Toward this unmet need, we have developed a dynamic constraint based modeling that can be used to interrogate cell free systems.

Genome scale stoichiometric reconstructions of microbial metabolism popularized by static, constraint based modeling techniques such as flux balance analysis (FBA) have become standard tools [147]. Since the first genome scale stoichiometric model of *E. coli*, developed by Edwards and Palsson [58], well over 100 organisms, including industrially important prokaryotes such as *E. coli* [67] or *B. subtilis* [183], are now available [70]. Stoichiometric models rely on a pseudo-steady-state assumption to reduce unidentifiable genome-scale kinetic models to an underdetermined linear algebraic system, which can be solved effi-

ciently even for large systems using linear programming. Traditionally, stoichiometric models have also neglected explicit descriptions of metabolic regulation and control mechanisms, instead opting to describe the choice of pathways by prescribing an objective function on metabolism. Interestingly, similar to early cybernetic models, the most common metabolic objective function has been the optimization of biomass formation [109], although other metabolic objectives have also been estimated [201]. Recent advances in constraint based modeling have overcome the early shortcomings of the platform, including describing metabolic regulation and control [106] and incorporating genome sequence into the model [3, 181]. Dynamic constraint based methods have also been developed in which the metabolic flux is computed over short-time intervals subject to time-varying constraints [162]. These methods are common, have been used in varied applications, [97, 96, 88, 93], and there are open source packages to support this class of calculation [100, 81, 82]. Thus, constraint based approaches, and their dynamic extensions, have proven useful in the discovery of metabolic engineering strategies and represent the state of the art in metabolic modeling [168, 265]. However, while constraint based tools have been used extensively to analyze whole cells systems, they have not yet been widely applied to study cell free reactions.

In this study, we constructed a dynamic constraint based model of cell-free protein expression. This approach avoids the pseudo-steady-state assumption found in traditional constraint based approaches, which allowed for the direct integration of dynamic metabolite measurements into the flux estimation problem, along with the accumulation or depletion of network metabolites. We adapted the sequence specific constraint based model of Vilkhovoy and coworkers [235] into a dynamic constraint based model of cell-free *E. coli*

metabolism and protein production, and leveraged the kinetic model of Horvath and coworkers [103] to provide synthetic data to inform metabolite constraints. CFPS synthesis is often (but not always) conducted in small scale batch reactors. Thus, the concentration of the components of the reaction mixture, and the associated rates of the metabolic processes in the reaction are not always constant. The Vilkhovoy et al study considered only the first hour of the CFPS reaction producing the model protein, chloramphenicol acetyltransferase (CAT). During this initial phase, the metabolic rates were approximately constant and a classical sequence specific flux balance analysis approach was sufficient to describe protein synthesis [235]. However, after this initial phase, there was a significant shift in productivity following the exhaustion of glucose (which occurred at approximately 1.5 hr). Horvath and coworkers developed a fully kinetic model that described the complete three hour reaction time course, including the shift in productivity following glucose exhaustion [103]. While this model described the CFPS dynamics, and the decrease in productivity following the exhaustion of glucose, the identification of the model from 37 measured metabolite trajectories was difficult. Thus, there was an unmet need for a tool that could describe the dynamics of a CFPS reaction, without the burden of identifying a full kinetic model. Toward this need, we developed a dynamic constraint based modeling approach for CFPS reactions which directly incorporated metabolite measurements (as constraints) into the flux calculation. The dynamic constraint based model satisfied time-dependent metabolite constraints, while predicting the concentration of the CAT protein and unconstrained metabolite concentrations. Model interrogation suggested the most important metabolite constraint was glucose, as excluding glucose yielded the greatest metabolite prediction error, and the greatest uncertainty in

the estimated metabolic flux. Further, we evaluated metabolite constraint sets with one more and one fewer metabolites than the base case (the 37 measured metabolites) to explore the impact of measurement selection on model performance. The single addition of metabolites yielded no significant improvement in the predictive power, while the single exclusion suggested glucose to be the most important measured metabolite in the base case. Next, we selected measurement species based on the results of singular value decomposition on the stoichiometric matrix. The top 36 species from the SVD analysis with the addition of glucose improved the predictive power and reduced flux uncertainty compared with the base case. Finally, we developed a heuristic optimization approach to estimate the optimal list of metabolite measurements. This approach significantly improved metabolite prediction compared to the base case. However, both the base measurement set and the heuristically optimized experimental design poorly characterized flux uncertainty. Taken together, this suggested dynamic constraint based modeling can aid in experimental design and measurement selection for metabolite prediction, but the flux estimation problem remains challenging. Further, while we modeled the cell-free production of only a single protein in this study, the sequence specific dynamic constraint based modeling approach presented here could be extended to multi-protein synthetic circuits, RNA circuits [104] or even small molecule production.

5.3 Results

5.3.1 Cell-free *E. coli* metabolic network.

We constructed the cell-free stoichiometric network by removing growth associated reactions from the *i*AF1260 reconstruction of K-12 MG1655 *E. coli* [67], and removing reactions not present in the cell-free system (see Materials and Methods). We then added the transcription and translation template reactions of Allen and Palsson for the CAT protein [3]. Thus, our stoichiometric network described the material and energetic demands for transcription and translation at sequence specific level. The metabolic network consisted of 264 reactions and 146 species; a schematic of the central carbon metabolism is shown in Fig. 5.1. The network described the major carbon and energy pathways and amino acid biosynthesis and degradation pathways. Lastly, we removed genes from the network that were knocked out in the *E. coli* host strain used to make the cell free extract (A19 ΔtonA ΔtnaA ΔspeA ΔendA ΔsdaA ΔsdaB ΔgshA), see Jewett et al for further details of the host strain and the cell free extract preparation [120]. Using this network, we simulated time-dependent cell-free production of the model protein CAT. We used dynamic modified flux balance analysis, a stoichiometric modeling technique that does not make the pseudo steady state assumption and allows the accumulation and depletion of metabolite species. Horvath and coworkers predicted time-dependent cell-free production of CAT using a fully kinetic model trained against an experimental dataset of 37 metabolites, including the substrate glucose, the protein product CAT, organic acids, amino acids, and energy species [103]. This model was used to generate the metabolite constraints used in this study. Transcription and translation rates were subject

to resource constraints encoded by the metabolic network, and transcription and translation model parameters were largely derived from literature (Table 5.1). In this study, we did not explicitly consider protein folding. However, the addition of chaperone or other protein maturation steps could easily be accommodated within the approach by updating the template reactions, see Palsson and coworkers [181]. The cell-free metabolic network and all model code and parameters can be downloaded under an MIT software license from the Varner-lab website [234].

5.3.2 Dynamic constrained simulation of cell free protein synthesis.

Cell free synthesis of the CAT protein showed two production phases, an initial fast production phase before glucose exhaustion (at approximately 1.5 hrs) and a slow production phase following glucose exhaustion. The metabolite profile varied significantly between these phases; for example, pyruvate and lactate were produced during the first phase but consumed during the second. Thus, a static pseudo steady state flux balance approach was not possible for this system. However, a central advantage of cell free systems is direct access to metabolite measurements, and the biosynthetic machinery during production. If we could directly integrate dynamic metabolite and protein concentration measurements into the flux estimation problem, we could potentially get a better estimate of the flux distribution. Toward this question, we developed a dynamic modeling approach in which metabolic fluxes were estimated so that all metabolites were non-negative and the simulated metabolites were

constrained to lie within a bounded range of the measured value. Using this technique, we simulated the cell free production of CAT subject to dynamic metabolite measurements.

We explored the influence of uncertainty in the transcription (TX) and translation parameters (TL) by sampling different values for the abundance and elongation rates of RNA polymerases and ribosomes, the polysome amplification constant, the mRNA degradation rate and other kinetic parameters appearing in the transcription and translation bounds. The base values for the TX/TL parameters are given in Table 5.1, and the uniform sampling procedure is described in the Materials and Methods. Central carbon metabolites (Fig. 5.2), amino acids (Fig. 5.3), and energy species (Fig. 5.4) in the synthetic measurement set were captured, within experimental error, by an ensemble of dynamic constraint based simulations. The 95% confidence interval (CI) estimates of the experimental error shown in each of the simulation figures was derived from the kinetic model of Horvath et al [103], which was trained on the 37 metabolites shown in Fig. 5.2 - 5.4. Thus, the 95% confidence estimates shown in this study were calculated based upon the kinetic model, which shows high fidelity with the experimental measurements. The flux estimation problem converged in greater than 99% of the simulation time intervals, given these metabolite constraints. This suggested there were not gross measurement errors in the measurement constraints, as the stoichiometric constraints were satisfied. Moreover, it suggested the error introduced by the time discretization scheme did not lead to inconsistent metabolite estimates. The ensemble of models captured the time evolution of protein biosynthesis, and the consumption and production of organic acid, amino acid and energy species. Arginine and glutamate were excluded from the constraint set, but were still largely captured by the en-

semble of dynamic constraint based models, although with wide variance than the synthetic measurement set. During the first hour, glucose was consumed as the primary carbon source for ATP, amino acids, and protein synthesis. After glucose was depleted, lactate and pyruvate were consumed as alternate substrates for energy production and CAT synthesis. Taken together, we captured the 37 metabolite measurements in the base synthetic data set, and captured the biphasic behavior of CAT production, although we significantly over-predicted the translation rate for some elements of the ensemble. This suggested there was excess capacity in the metabolic network, that could be used to enhance protein production.

We quantified the uncertainty in the estimated metabolic flux distribution, given constrained CAT production using flux variability analysis (FVA) for the base synthetic data across the three hours of measurement (Tables 5.2 and 5.3). The analysis was divided into two phases: phase 1 where glucose was consumed as the carbon source, and phase 2 when glucose was depleted and lactate and pyruvate were utilized. The reactions associated with protein synthesis (translation initiation, translation, tRNA charging, mRNA degradation) were unsurprisingly the most constrained, as CAT production was forced to remain the same. Transcription was not varied in this analysis. On the other hand, glycolytic, pentose phosphate, and Entner-Doudoroff reactions were not highly constrained, indicating the robustness of substrate utilization. However, one exception to this was the net reaction through *zwf* reaction, which was tightly constrained, suggesting that glycolysis alone cannot support protein production. Interestingly, although the two phases consumed different carbon sources, the flux variability remained similar. Taken together, these results suggested there was significant flexibility in the ability of the metabolic network to meet

the carbon and energy demands of protein synthesis. Next, we explored alternative measurement sets to constrain the simulation of cell free protein synthesis.

5.3.3 Alternative measurement sets.

The base synthetic data set, consisting of 36 metabolite time series and the protein product CAT, was the measurement set used to train the kinetic model of Horvath et al [103]. Thus, the error confidence intervals used in this study as constraints on the flux estimation problem were informed by experimental measurements of glucose, organic and amino acids, energy species and the protein product CAT. However, we have no a priori reason to suppose that this experimental design was optimal. Toward this question, we performed simulations and flux variability analysis (FVA) for alternative synthetic data sets to understand the importance of measurement selection when characterizing CFPS (Fig. 5.5 and Fig. 5.6). In all cases, we assumed the same sampling frequency as the base synthetic dataset, but we varied which species were measured. First, we removed each of the 37 metabolites from the base set, one at a time, to create 37 measurement exclusion sets, consisting of 36 metabolites each (Fig. 5.5, light gray dots). For each set, the state the dynamic model was used to calculate a value of error against the synthetic data, and FVA was used to calculate a value of flux uncertainty. Most of the exclusion sets clustered around the base case, with error values between 75% and 110%, and flux uncertainties between 93% and 103%, of the base case. The exception to this was the glucose exclusion set, which showed 89% higher error and 7% greater flux uncertainty. Within the primary cluster, a slight pattern emerged: the sets in which an organic acid was removed tended to result in increased error, while the removal of an amino acid

tended to reduce error; however, this was not true across all metabolites. We also performed the analysis on several inclusion sets to determine which additional metabolites could improve predictive power (Fig. 5.5, black dots). In particular, we added unmeasured central carbon metabolites to the base case which resulted in 23 inclusion sets, consisting of 38 metabolites each. As with the exclusion sets, most of the inclusion sets clustered around the base case, with error values between 72% and 103%, and flux uncertainties between 94% and 102%, of the base case. Considering all exclusion and inclusion sets, there was generally no correlation between the metabolite prediction error and flux uncertainty. Taken together, these suggested central carbon metabolites, especially glucose, were important to characterizing the network, but performing single additional measurements was not enough to significantly increase predictive power. Next, we explored whether measurement selection could be based upon the structural features of a metabolic network. Toward this question, we used singular value decomposition (SVD) of the stoichiometric matrix to suggest which metabolites should be measured.

Singular value decomposition (SVD) measurement selection outperformed the base case, with a prediction error improvement of 11% and similar flux variability (Fig. 5.5, open square). SVD was used to decompose the stoichiometric matrix into 105 modes. The top 36 metabolites that had the greatest weighted sum across the modes that accounted for 95% of the network were estimated. Since our exclusion analysis identified glucose as the single most important metabolite, we added it to the top metabolites as determined by SVD to obtain a 37-metabolite constraint set, consisting of: GTP, GDP, GMP, ATP, ADP, AMP, UTP, UMP, CTP, CMP, GLN, GLU, ASP, LYS, LEU, HIS, THR, PHE, ALA, VAL, TYR, GLY, SER, H, ASN, ILE, MET, AKG, PYR, ARG, CYS, NH3, FUM, SUCC,

TRP, ACCOA, and glucose. The SVD measurement set suggested that energy, and amino acid species carried the most information compared to central carbon species which made up a relatively smaller fraction of the list. Surprisingly, the measurement set selected by SVD was approximately 80% similar to the original synthetic data generated by hand. However, the 20% difference was enough to improve the prediction error by approximately 11%. Taken together, measurements selected by SVD decomposition of the stoichiometric matrix improved the prediction of metabolite abundance, but SVD-based measurement selection did not improve flux variability.

Next, we used heuristic optimization to systematically investigate the effect of changing the dimension and identify of the measurement constraints (Fig. 5.6). In particular, we minimized the error and flux variability of model predictions by varying the metabolites which appeared in the synthetic constraint set. We used a binary simulated annealing algorithm to switch metabolite membership in the constraint set on or off, and thus generated an ensemble of > 200 measurement constraint sets (Fig. 5.6A). While there was no strict error threshold, the simulated annealing algorithm was less likely to accept high-error sets into the ensemble; thus, the error of most sets in the ensemble was less than that of the base case. Specifically, the error varied from just over double to less than one ten-thousandth of the base case. Flux uncertainty was also a component of the objective function, but was only improved by 7%, suggesting this performance metric was tightly constrained; network flux values were not be well characterized, even with comprehensive training datasets. As expected, there was an inverse relationship between the number of metabolite constraints and the prediction error (Fig. 5.6B). However, the slope of that trend was striking; error was improved by three to four orders of magnitude, simply

by increasing the number of constraints by 11 or fewer. Furthermore, the base synthetic measurement set was outperformed by the majority of the ensemble; often the simulated annealing approach achieved the same error with fewer constraints, or much lower error with the same, or even fewer, constraints. This suggests that while comprehensive, the original synthetic dataset was not optimal in terms of predictive power per measurement. However, the base case was one of the best in terms of reducing flux uncertainty.

Lastly, we investigated which metabolites were most effective at improving predictive power by considering how often it appeared in the ensemble (Table 5.4). Glucose unsurprisingly appeared most often (tied with G6P), but interestingly was not in every constraint set. Those that did not contain glucose had some of the highest errors, but also some of the smallest constraint set sizes (Fig. 5.6B, black dots). The most frequent metabolites from the heuristic method were largely from glycolysis, pentose phosphate, and the TCA cycle (compared to the SVD analysis which gave greater consideration to energetic and amino acid species). To further understand the species selection, we calculated the frequency of appearance in the 57 best sets, those with error of least three orders of magnitude lower than the base case (Table 5.5). Nineteen metabolites appeared in all of these sets, and all but Alanine were central carbon metabolites (defined here as glycolysis, pentose phosphate, TCA). Taken together, measurement selection made a significant difference in capturing dynamic metabolite abundance in cell free protein synthesis. Although the error decreased with increasing measurement number overall, the specific combination of metabolites was arguably even more important. Metabolic fluxes, however, remained unknown despite the large number of measurements taken.

5.4 Discussion

In this study, we presented a dynamic constraint based model of cell-free protein expression. This approach avoids the pseudo-steady-state assumption found in traditional constraint based approaches, which allowed for the direct integration of metabolite measurements into the flux estimation problem, and the the accumulation or depletion of network metabolites. The approach used the *E.coli* cell free protein synthesis metabolic network from Vilkhovoy and coworkers [235], and the simulated metabolite trajectories from the kinetic model of Horvath et al [103] as constraints on the CFPS flux calculation. The dynamic constraint based model satisfied time-dependent metabolite measurement constraints, predicted unconstrained metabolite concentrations as well as the concentration of a model protein, chloramphenicol acetyltransferase (CAT). Model interrogation suggested the most important metabolite measurement within the dataset to be glucose, as excluding the glucose yielded the greatest metabolite prediction error, and the greatest uncertainty in the estimated metabolic flux. Further, we evaluated metabolite constraint sets with one more and one fewer metabolites than the base case (37 metabolites) to explore the impact of measurement selection on model performance. The single addition of metabolites yielded no significant improvement in the predictive power, while the single exclusion suggested glucose to be the most important measured metabolite in the base case. Next, we selected measurement species based on the results of singular value decomposition on the stoichiometric matrix. The top 36 species from the SVD analysis with the addition of glucose improved the predictive power and reduced flux uncertainty compared with the base case. Finally, we described a heuristic optimization approach to estimate the optimal list of

metabolite measurements. Measurement sets determined by heuristic optimization vastly outperformed the accuracy of the base synthetic dataset; model precision, meanwhile, was virtually unchanged despite comprehensive measurement sets. Taken together, model interrogation showed that even with a comprehensive dataset, there still exists a great amount of uncertainty associated with metabolic fluxes. This highlights the need for fluxomic data to fully understand biological networks.

Despite synthetic datasets consisting of greater than 30 metabolite time series, estimates of metabolic flux were largely uncertain. Flux variability analysis suggested that the metabolite constraints could be met with a wide range of different flux distributions. For instance, an open question in cell-free systems is the balance between glycolytic versus pentose phosphate pathway flux. In previous studies of *E. coli* cell free protein synthesis, the kinetic model of Horvath and coworkers suggested that glucose was consumed primarily by glycolytic reactions, with minimal flux into the pentose phosphate pathway. However, Vilkhovoy et al estimated, using sequence specific flux balance analysis with the same experimental dataset, that the CAT production was unaffected by the choice of pentose phosphate pathway versus glycolysis; deletion of either pathway did not change protein productivity. To answer this discrepancy, model analysis showed, during the first phase when glucose was being consumed, glycolytic and pentose phosphate fluxes (pgi and zwf, respectively) exhibited large uncertainty, as either could be utilized to satisfy CAT production. The measurement selection analysis was conducted by excluding or including a metabolite from the constraint set. The exclusion sets were dominated by the removal of glucose, and to a lesser extent the organic acids, suggesting measurements of central carbon metabolism intermediates were more important than energetic

and amino acid measurements. However, the inclusion sets showed no significant effect on error and flux uncertainty. There was generally no correlation between the error and flux uncertainty of a model constrained to a particular metabolite set, except with respect to the outlier glucose. Model calculations showed that even with a comprehensive data set of 37 metabolite measurements, there was significant flux uncertainty. This suggested there were many flux combinations that could give rise to the same set of time course measurements. This phenomenon was further supported by analyzing the ensemble of constraint sets determined by heuristic optimization. Although the optimization algorithm reduced the objective function by four degrees of magnitude, the flux variability remained stagnant in comparison. An ensemble of measurement sets ranging from 22 to 48 metabolite constraints was only able to reduce flux uncertainty by 7% from the base synthetic data set. The dynamic constraint based model showed high flux variability in important branch points, including the glucose-6-phosphate split between glycolysis and pentose phosphate, the 6PGC split into pentose phosphate and Entner-Doudoroff, and the pyruvate split into TCA cycle versus lactate production. This may be why the high overall flux variability was robust to the varying of metabolite constraints. Using three different sampling approaches (single additions/exclusions, singular value decomposition, simulated annealing) coupled with the dynamic constraint based model, we estimated key metabolites that could be prioritized in measurement selection, such as glucose. Although measuring central carbon metabolites and amino acids is the intuition of most researchers, model interrogation was able to provide the importance of certain species over others; for instance, measuring G6P, G3P, and F6P would be more fruitful than measuring PEP. Interestingly, many of the most valuable measurements were involved in upper gly-

colysis and pentose phosphate, such as glucose, G6P, and 6PGL. This may be because upstream metabolites have an effect on more of the network; any error or uncertainty in these metabolites will cascade down the rest of the network and magnify throughout. Taken together, the dynamic constraint based model quantitatively affirmed the robustness of metabolism, and illustrated the complexity of inferring flux information from metabolite concentrations. Ultimately, to determine the metabolic flux distribution occurring in a cell free system, we need to add additional constraints to the flux estimation calculation. This study suggested metabolite measurements alone were not sufficient. However, these are not the only experimentally realizable types of constraints. For example, thermodynamic feasibility constraints may result in a better depiction of the flux distribution [92, 85], and ^{13}C labeling constraints in cell free systems could provide significant insight. However, while ^{13}C labeling techniques are well established for *in vivo* processes [258], application of these techniques to cell free systems remains an active area of research.

In summary, we used a dynamic constraint based modeling approach to simulate cell free metabolism, and to study how measurement selection impacts model performance. We extended sequence specific flux balance analysis, by removing the pseudo steady state assumption, and adding synthetic metabolite measurement constraints to the flux calculation. Using this method, we simulated the cell free synthesis of a model protein, chloramphenicol acetyltransferase, we identified the most important measured species in the cell free system, and additional species that yielded the lowest metabolite prediction error and flux uncertainty. Only synthetic metabolite measurements were used in this study; however, this work built a foundation to rationally design experimental measurement protocols, that could be implemented with a variety

of analytical techniques. Taken together, these findings represent a novel tool for dynamic cell free simulations, measurement selection and pathway analysis, not only for *E. coli*, but potentially for align variety of metabolic networks, whether *in vivo* or cell-free. However, while this first study was promising, there were several issues to consider in future work. First, while we described transcription and translation at a sequence specific level, we have not considered the complexities of protein folding, or post-translational modifications such as protein glycosylation. A more detailed description of transcription and translation reactions, including the role of chaperones in protein folding, has been used in *in-vivo* genome scale ME models e.g., see O'Brien et al [181]. These template reactions could easily be adapted to a cell free system, thereby providing a potentially higher fidelity description protein synthesis and folding. Next, the inclusion of post-translational modifications such as protein glycosylation in the next generation of models will be important to describe the cell free synthesis of therapeutic proteins. DeLisa and coworkers recently showed that glycoproteins can be synthesized in a cell free system, using extract generated from modified *E.coli* cells capable of asparagine-linked protein glycosylation [113]. Simulation of the generation and attachment of glycans to protein targets could be an important step to optimizing cell free glycoprotein production. Lastly, while we modeled the cell-free production of a only single protein in this study, sequence specific dynamic constraint models could be developed for multi-protein synthetic circuits, RNA circuits or even small molecule production. Thus, this approach offers a unique tool to model and potentially optimize a wide variety of application areas in synthetic biology.

5.5 Materials and Methods

5.5.1 Formulation of the model equations.

We modeled the time evolution of the ith metabolite concentration (x_i), the scaled activity of network enzymes (ϵ_i), transcription processes generating the mRNA m and translation processes generating the protein $\mathcal{P}$ in an *E. coli* cell-free metabolic network as a system of ordinary differential equations:

$$\dot{x}_i = \sum_{j=1}^{\mathcal{R}} \sigma_{ij} r_j(\mathbf{x}, \epsilon, \mathbf{k}) \qquad i = 1, 2, \ldots, \mathcal{M} \tag{5.1}$$

$$\dot{\epsilon}_i = -\lambda_i \epsilon_i \qquad i = 1, 2, \ldots, \mathcal{N} \tag{5.2}$$

$$\dot{m} = r_T - \lambda m \tag{5.3}$$

$$\dot{\mathcal{P}} = r_X \tag{5.4}$$

The quantity $\mathcal{R}$ denotes the number of metabolic reactions, $\mathcal{M}$ denotes the number of metabolites and $\mathcal{N}$ denotes the number of metabolic enzymes in the model. The quantity $r_j(\mathbf{x}, \epsilon, \mathbf{k})$ denotes the rate of reaction j. Typically, reaction j is a non-linear function of metabolite and enzyme abundance, as well as unknown kinetic parameters $\mathbf{k}$ ($\mathcal{K}\times 1$). The quantity σ_{ij} denotes the stoichiometric coefficient for species i in reaction j. If $\sigma_{ij} > 0$, metabolite i is produced by reaction j. Conversely, if $\sigma_{ij} < 0$, metabolite i is consumed by reaction j, while $\sigma_{ij} = 0$ indicates metabolite i is not connected with reaction j. Lastly, λ_i denotes the scaled enzyme activity decay constant. The system material balances were subject to the initial conditions $\mathbf{x}(t_o) = \mathbf{x}_o$ and $\epsilon(t_o) = \mathbf{1}$ (initially we have 100% cell-free enzyme activity).

The cell-free model equations were solved using a dynamic constraint based approach in which the rates of the metabolic fluxes, transcription and translation processes were estimated by solving an optimization subproblem from t to $t + \Delta t$. In particular, the biochemical fluxes $r_1, r_2, \ldots, r_{\mathcal{R}}$ which appear in the balance equations were calculated from t to $t + \Delta t$ by solving a constrained optimization subproblem with (potentially non-linear) objective $O(x_1, x_2, \ldots, x_{\mathcal{M}})$:

$$\max_{r_1, r_2, \ldots, r_{\mathcal{R}}} O(x_1, x_2, \ldots, x_{\mathcal{M}}) \tag{5.5}$$

subject to species constraints and flux bounds:

$$\left(\sum_{j=1}^{\mathcal{R}} \sigma_{ij} r_j - \dot{x}_i \right) \geq 0 \qquad i = 1, 2, \ldots, \mathcal{M} \tag{5.6}$$

$$0 \leq r_j \leq \mathcal{U}_j(x_1, x_2, \ldots, x_{\mathcal{M}}, \kappa) \qquad j = 1, 2, \ldots, \mathcal{R} \tag{5.7}$$

In this study, we maximized the rate of translation r_X unless otherwise specified. We discretized the derivative term for each species using a constant width h forward different approximation (however, this was done for convenience and more sophisticated techniques could have been used). The reaction bounds $\mathcal{U}_j(x_1, x_2, \ldots, x_{\mathcal{M}}, \kappa)$ are *potentially* complex functions of the system state, and can be updated during each time step. Here, we modeled the upper bound for flux j as $\hat{V}_{max}\epsilon_j(k)$, where $\hat{V}_{max}$ denotes a characteristic maximum reaction velocity, and $\epsilon_j(k)$ denotes the scaled enzyme activity catalyzing reaction j at time step k. The characteristic maximum reaction velocity was set to 600 mM/hr (which corresponds to an average $k_{cat} \simeq 1000\text{s}^{-1}$ and and enzyme concentration of approximately 0.2 μM) unless otherwise specified. Additional species constraints can be added to *directly* incorporate metabolomic, proteomic or transcriptomic

measurements into the flux calculation. In this study, we incorporated metabolite measurement constraints of the form:

$$\chi^{L}_{m,k+1} \leq x_{m,k+1} \leq \chi^{U}_{m,k+1} \qquad m = 1, 2, \ldots, \Xi \tag{5.8}$$

where $\chi^{L}_{m,k+1}$ and $\chi^{U}_{m,k+1}$ denote the lower and upper measurement bound for metabolite m at time step $k + 1$, where Ξ metabolites were measured over the time course of the cell free reaction. Lastly, we imposed a user-configurable bound $\mathcal{B}_i$ on the maximum rate of change for metabolite i:

$$|\dot{x}_i| \leq \mathcal{B}_i \qquad i = 1, 2, \ldots, \mathcal{M} \tag{5.9}$$

and non-negativity constrains $x_i \geq 0$ for all metabolites and all time steps.

The bounds on the transcription rate ($\mathcal{L}_T = r_T = \mathcal{U}_T$) were modeled as:

$$r_T = V_T^{max}\left(\frac{G_{\mathcal{P}}}{K_T + G_{\mathcal{P}}}\right) \tag{5.10}$$

where $G_{\mathcal{P}}$ denotes the concentration of the gene encoding the protein of interest, and K_T denotes a transcription saturation coefficient. The maximum transcription rate V_T^{max} was formulated as:

$$V_T^{max} \equiv \left[R_T\left(\frac{\dot{v}_T}{l_G}\right)u(\kappa)\right] \tag{5.11}$$

where R_T denotes the RNA polymerase concentration (nM), $\dot{v}_T$ denotes the RNA polymerase elongation rate (nt/h), l_G denotes the gene length (nt). The term $u(\kappa)$ (dimensionless, $0 \leq u(\kappa) \leq 1$) is an effective model of promoter activity, where κ

denotes promoter specific parameters. The general form for the promoter models was taken from Moon *et al.* [174]; which was based on earlier studies from Bintu and coworkers [13], and similar to the genetically structured modeling approach of Lee and Bailey [143]. In this study, we considered only the T7 promoter model:

$$u_{T7} = \frac{K_{T7}}{1 + K_{T7}} \tag{5.12}$$

where K_{T7} denotes a T7 RNA polymerase binding constant. The values for all promoter parameters are given in Table 5.1.

The translation rate (r_X) was bounded by:

$$0 \leq r_X \leq V_X^{max}\left(\frac{m}{K_X + m}\right) \tag{5.13}$$

where m denotes the mRNA abundance and K_X denotes a translation saturation constant. The maximum translation rate V_X^{max} was formulated as:

$$V_X^{max} \equiv \left[K_P R_X\left(\frac{\dot{v}_X}{l_P}\right)\right] \tag{5.14}$$

The term K_P denotes the polysome amplification constant, $\dot{v}_X$ denotes the ribosome elongation rate (amino acids per hour), and l_P denotes the number of amino acids in the protein of interest. The mRNA abundance m was estimated as:

$$m_{k+1} = m_k + (r_T - m_k\lambda)h \tag{5.15}$$

where λ denotes the mRNA degradation rate constant (h^{-1}). All translation parameters are given in Table 5.1.

Sampling of transcription and translation parameters.

The influence of the uncertainty in the transcription (TX) and translation (TL) parameters was estimated by sampling the expected physiological ranges for these parameters as determined from literature. We generated uniform random samples between an upper (u) and lower (l) parameter bound of the form:

$$p^* = l + (u - l) \times \mathcal{U}(0, 1) \tag{5.16}$$

The T7 RNA polymerase concentration was sampled between 990 and 1010 nM, ribosome levels between 1.8 and 2.2 μM, the RNA polymerase elongation rate between 20 and 30 nt/s, and the ribosome elongation rate between 1.0 and 3.0 aa/s [228, 79].

5.5.2 Generation and evaluation of alternative measurement sets.

The measurement sets consisted of the base (one set of 37 metabolites), inclusion sets (23 sets of 38 metabolites each), exclusion sets (37 sets of 36 metabolites each), SVD-guided (one set of 37 metabolites), and simulated annealing samples (238 sets of varying length). In all cases, we assumed the same sampling frequency as the base synthetic dataset, but we varied which species were measured. The exclusion or inclusion measurement sets were constructed by removing or adding a metabolite to the base set, while the SVD-guided measurement set was constructed from high importance metabolites; the top 36 metabolites (plus glucose) that had the greatest singular value weighted sum across

the SVD-modes, accounting for 95% of the network structure, were designated the SVD measurement set. Lastly, we used simulated annealing to generate potentially optimal measurements sets, where the objective was to minimize the product of the prediction error, and flux uncertainty. The prediction error, $\mathcal{E}$, was computed by comparing the simulated versus the measured value of a metabolite, for a $\mathcal{M}_{\text{core}}$ set of metabolites. On the other hand, the flux variability was computed using flux variability analysis (FVA) [176], subject to constraints on the CAT production rate, and the selected metabolite trajectories. In particular, the metabolite prediction error was calculated from the time-dependent state array:

$$\mathcal{E} = \sum_{i=1}^{\mathcal{M}_{\text{core}}} \sum_{t=t_i}^{\mathcal{T}} \left(\max\left(x_i(t) - y_i^U(t), 0\right) + \max\left(y_i^L(t) - x_i(t), 0\right) \right)$$

where $x_i(t)$ denotes the simulated value of metabolite i at time t, $y_i^U(t)$ denotes the upper bound of the 95% confidence interval on the synthetic data for metabolite i at time t, $y_i^L(t)$ denotes the lower bound of the 95% confidence interval on the synthetic data for metabolite i at time t, and $\mathcal{M}_{\text{core}}$ denotes the subset of metabolites in the core metabolism. For this calculation, the entire time course was considered (t_i = 0 h, $\mathcal{T}$ = 3 h). The flux uncertainty was calculated from the maximal and minimal flux arrays:

$$\sigma_{\text{overall}} = \sum_{r_j \in \mathcal{R}_{\text{core}}} \sum_{t=t_i}^{\mathcal{T}} \left(r_j^{\max}(t) - r_j^{\min}(t) \right)^2$$

where $r_j^{\max}(t)$ denotes the maximum value of flux j, while $r_j^{\min}(t)$ denotes the value of flux j at time t, calculated using flux variability analysis. The quantity $\mathcal{R}_{\text{core}}$ denotes the subset of reactions that constitute the core metabolism. For the flux uncertainty calculations, either the entire reaction time course was considered (t_i = 0 h, $\mathcal{T}$ = 3 h), or the uncertainty was calculated separately for each phase (phase 1: t_i = 0 h, $\mathcal{T}$ = 1 h; phase 2: t_i = 1 h, $\mathcal{T}$ = 3 h).

The simulated annealing algorithm began by evaluating the error and flux uncertainty of the base case and multiplying these to obtain a cost function:

$$\texttt{cost} = \mathcal{E} \cdot \sigma_{\text{overall}} \tag{5.17}$$

Then, each metabolite that was considered measurable was added to or removed from the constraint set with a certain probability p_{switch}:

$$\theta_i^{\text{new}} = \begin{cases} 1 - \theta_i & \mathcal{R}_{0,1}^{\text{uniform}} < p_{switch} \\ \theta_i & \mathcal{R}_{0,1}^{\text{uniform}} > p_{switch} \end{cases} \qquad i = 1, 2, \ldots, \mathcal{M}_{\text{measurable}} \tag{5.18}$$

where $\theta_i \in \{0, 1\}$ denotes a binary parameter encoding whether or not metabolite i is in the constraint set, $\mathcal{R}_{0,1}^{\text{uniform}}$ denotes a uniform random number taken from a distribution between 0 and 1, and $\mathcal{M}_{\text{measurable}}$ denotes the set of metabolites deemed to be measurable. For each newly generated constraint set, we re-solved the dFBA and FVA problems, and re-calculated the cost function. All sets with a lower cost were accepted into the ensemble. Sets with a higher cost were also accepted into the ensemble, if they satisfied the acceptance constraint:

$$\mathcal{R}_{0,1}^{\text{uniform}} < exp\left(-\alpha \cdot \frac{\texttt{cost}_{\texttt{new}} - \texttt{cost}}{\texttt{cost}}\right) \tag{5.19}$$

where $\mathcal{R}_{0,1}^{\text{uniform}}$ denotes a random number taken from a uniform distribution between 0 and 1, $\texttt{cost}$ denotes the cost of the current parameter set, $\texttt{cost}_{\texttt{new}}$ denotes the cost of the new parameter set, and α denotes an adjustable parameter to control the tolerance to high-error sets. A total of 238 samples were accepted into the ensemble, of which there were 219 unique sets. Both $\mathcal{M}_{\text{core}}$ and $\mathcal{R}_{\text{core}}$ and user-configurable, and defined in the model code repository available from the Varnerlab website [234].

This study was supported by a National Science Foundation Graduate Research Fellowship (DGE-1333468) to N.H. Research reported in this publication was also supported by the Systems Biology Coagulopathy of Trauma Program with support from the US Army Medical Research and Materiel Command under award number W911NF-10-1-0376.

J.V directed the modeling study. D.D, N.H and J.V developed the cell-free protein synthesis mathematical model, and parameter ensemble. Simulations were conducted by D.D and N.H. The manuscript was prepared and edited for publication by D.D, N.H, and J.V.

The authors declare no conflict of interest. The funding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

Table 5.1: Reference values for transcription, translation, and mRNA degradation from literature. Transcription rate calculated from elongation rate, mRNA length, and promoter activity level. Translation rate calculated from elongation rate, protein length, and polysome amplification constant. The mRNA degradation rate calculated from a characteristic mRNA half-life.

Description	**Parameter**	**Value**	**Units**	**Reference**
T7 RNA polymerase concentration	R_T	1.0	μM	
Ribosome concentration	R_X	2	μM	[79]
CAT mRNA length	l_G	660	nt	[133]
CAT protein length	l_P	219	aa	[133]
Transcription saturation coefficient	K_T	100	nM	estimated
Transcription elongation rate	$\dot{v}_T$	25	nt/s	[79]
Translation saturation coefficient	K_X	45	μM	estimated
Translation elongation rate	$\dot{v}_X$	1.5	aa/s	[79]
T7 Promoter activity level	u	0.9		estimated
Transcription rate	$k_{cat}^T = \left(\frac{\dot{v}_T}{l_G}\right) u$	123	h^{-1}	calculated
Polysome amplification constant	K_P	10		estimated
Translation rate	$k_{cat}^X = \left(\frac{\dot{v}_X}{l_P}\right) K_P$	247	h^{-1}	calculated

mRNA degradation time	$t_{1/2}$	8	min	BNID 106253

Table 5.2: Flux uncertainty calculated using flux variability analysis for the base synthetic dataset during the first production phase (0 h to 1.5 h), normalized to the glucose consumption rate.

Effector	Effect	Target
Source		
RNA polymerase	Translation	<0.01
RNA polymerase	Translation initiation	<0.01
tRNA charging of alanine	tRNA charging (ALA)	<0.01
tRNA charging of cysteine	tRNA charging (CYS)	<0.01
tRNA charging of aspartate	tRNA charging (ASP)	<0.01
tRNA charging of histidine	tRNA charging (HIS)	<0.01
tRNA charging of serine	tRNA charging (SER)	<0.01
tRNA charging of tyrosine	tRNA charging (TYR)	<0.01
tRNA charging of phenylalanine	tRNA charging (PHE)	<0.01
tRNA charging of arginine	tRNA charging (ARG)	<0.01
tRNA charging of glutamate	tRNA charging (GLU)	<0.01
mRNA degradation	mRNA degradation	<0.01
tRNA charging of tryptophan	tRNA charging (TRP)	<0.01
tRNA charging of proline	tRNA charging (PRO)	<0.01
tRNA charging of asparagine	tRNA charging (ASN)	<0.01
tRNA charging of isoleucine	tRNA charging (ILE)	<0.01
tRNA charging of glycine	tRNA charging (GLY)	<0.01
tRNA charging of glutamine	tRNA charging (GLN)	<0.01
tRNA charging of lysine	tRNA charging (LYS)	<0.01
tRNA charging of threonine	tRNA charging (THR)	<0.01
tRNA charging of valine	tRNA charging (VAL)	<0.01
tRNA charging of methionine	tRNA charging (MET)	<0.01
tRNA charging of leucine	tRNA charging (LEU)	<0.01
Step 6 of AMP synthesis	R_A_syn_6	<0.01
Orotate synthase 1	R_or_syn_1	<0.01
Metionine biosynthesis	R_met	0.01
Valine biosynthesis	R_val	0.01

Continued on next page

Enzyme/Pathway	Reaction	Uncertainty
Leucine biosynthesis	R_leu	0.01
Aldhyde-alcohol dehydrogenase	R_adhE_net	0.01
Malate dehydrogenase	R_mdh_net	0.01
Glycine biosynthesis	R_gly_deg	0.02
Threonine degradation 2	R_thr_deg2	0.02
Acetate kinase	R_ackA_net	0.02
Alanine biosynthesis	R_alaAC_net	0.02
Isoleucine biosynthesis	R_ile	0.03
Tyrosine biosynthesis	R_tyr	0.03
Histidine biosynthesis	R_his	0.03
Methylglyoxal degradation	R_mglx_deg	0.03
Transaldolase	R_talAB_net	0.04
Glycine cleavage system	R_gly_fol_net	0.04
Ribulose-phosphate 3-epimerase	R_rpe_net	0.04
Phosphate acetyltransferase	R_pta_net	0.04
Phosphoglycerate kinase	R_pgk_net	0.05
Glyceraldehyde-3-phosphate dehydrogenase	R_gapA_net	0.05
Fructose 1,6-bisphosphate aldolase	R_fbaA_net	0.05
Enolase	R_eno_net	0.05
Phenylalanine biosynthesis	R_phe	0.05
Transketolase 2	R_tkt2_net	0.06
Fumarate hydratase	R_fum_net	0.06
Transketolase 1	R_tkt1_net	0.07
Orotate synthase 2	R_or_syn_2	0.07
Phosphoglycerate mutase	R_gpm_net	0.08
Ribose-5-phosphate isomerase	R_rpi_net	0.09
CTP synthetase 1	R_ctp_1	0.09
CTP synthetase 2	R_ctp_2	0.09
Triosephosphate isomerase	R_tpiA_net	0.1
Step 7 of AMP synthesis	R_A_syn_7	0.15
Step 12 of AMP synthesis	R_A_syn_12	0.17

Continued on next page

Enzyme/Pathway	Reaction	Uncertainty
Lactate dehydrogenase	R_ldh_net	0.17
Step 5 of AMP synthesis	R_A_syn_5	0.2
Methylenetetrahydrofolate reductase	R_mthfr2a	0.2
Glucose-6-phosphate dehydrogenase	R_zwf_net	0.21
Methylenetetrahydrofolate dehydrogenase	R_mthfd_net	0.21
UMP synthesis	R_ump_syn	0.22
OMP synthesis	R_omp_syn	0.22
Lysine degradation	R_lys_deg	0.23
Lysine biosynthesis	R_lys	0.23
Isocitrate dehydrogenase	R_icd_net	0.23
Threonine degradation 3	R_thr_deg3	0.24
Step 8 of AMP synthesis	R_A_syn_8	0.26
Tryptophan degradation	R_trp_deg	0.27
Methylenetetrahydrofolate dehydrogenase	R_mthfc_net	0.28
Tryptophan biosynthesis	R_trp	0.28
Aconitase	R_acn_net	0.33
Phosphoglucose isomerase	R_pgi_net	0.33
Step e of folate synthesis	R_fol_e	0.34
Step 4 of AMP synthesis	R_A_syn_4	0.4
GMP synthetase	R_gmp_syn	0.44
Step 9 of AMP synthesis	R_A_syn_9	0.48
XMP synthase	R_xmp_syn	0.53
Step 3 of AMP synthesis	R_A_syn_3	0.6
Step 10 of AMP synthesis	R_A_syn_10	0.63
Step 2b of folate synthesis	R_fol_2b	0.63
Glutamate dehydrogenase	R_gdhA_net	0.71
Step 3 of folate synthesis	R_fol_3	0.74
Step 4 of folate synthesis	R_fol_4	0.79
Pyruvate formate lyase	R_pflAB	0.8
Step 2 of AMP synthesis	R_A_syn_2	0.81
Step 2a of folate synthesis	R_fol_2a	0.81

Continued on next page

Enzyme/Pathway	Reaction	Uncertainty
Step 1 of folate synthesis	R_fol_1	0.99
Glucokinase	R_glk_atp	1
Step 1 of AMP synthesis	R_A_syn_1	1
Arginine degradation	R_arg_deg	1.23
Glycine biosynthesis	R_glyA	1.33
Phosphoribosylpyrophosphate synthase	R_prpp_syn	1.34
Chorismate synthesis	R_chor	1.35
Succinate thiokinase	R_sucCD	1.55
2-Ketoglutarate dehydrogenase	R_sucAB	1.55
GABA degradation 1	R_gaba_deg1	1.56
GABA degradation 2	R_gaba_deg2	1.56
Glutamate degradation	R_glu_deg	1.56
Arginine biosynthesis	R_arg	1.68
Pyruvate dehydrogenase	R_pdh	2.06
Malate synthase	R_aceB	2.3
Threonine degradation 1	R_thr_deg1	2.32
Isocitrate lyase	R_aceA	2.36
Threonine biosynthesis	R_thr	2.48
Citrate synthase	R_gltA	2.62
6-Phosphogluconate dehydrogenase	R_gnd	2.62
Cysteine biosynthesis	R_cysEMK	4.59
Cysteine degradation	R_cys_deg	4.6
Proline biosynthesis	R_pro	5.45
Proline degradation	R_pro_deg	5.47
6-Phosphogluconate dehydrase	R_edd	5.96
2-Keto-3-deoxy-6-phospho-gluconate aldolase	R_eda	5.96
Serine degradation	R_ser_deg	6.43
Nucleotide diphosphatase (ATP)	R_atp_amp	6.53
Nucleotide diphosphatase (UTP)	R_utp_ump	6.53
Nucleotide diphosphatase (GTP)	R_gtp_gmp	6.53
Nucleotide diphosphatase (CTP)	R_ctp_cmp	6.53

Continued on next page

Enzyme/Pathway	Reaction	Uncertainty
Cytidylate kinase	R_atp_cmp	6.56
Guanylate kinase	R_atp_gmp	6.6
UMP kinase	R_atp_ump	6.62
6-Phosphogluconolactonase	R_pgl	6.67
Serine biosynthesis	R_serABC	6.72
NADH:ubiquinone oxidoreductase	R_nuo	7.39
NADH dehydrogenase 1	R_ndh1	7.39
NADH dehydrogenase 2	R_ndh2	7.39
Fumurate reductase	R_frd	7.44
Succinate dehydrogenase	R_sdh	7.93
Malic enzyme A	R_maeA	7.99
Malic enzyme B	R_maeB	8.01
Cytochrome oxidase bo	R_cyo	8.03
Cytochrome oxidase bd	R_cyd	8.03
ATP synthase	R_atp	11.71
PEP synthase	R_pps	12.98
Fructose-1,6-bisphosphate aldolase	R_fdp	12.98
Adenosinetriphosphatase	R_atp_adp	12.98
PEP carboxykinase	R_pck	12.98
Asparagine biosynthesis	R_asnB	13
Glutamate biosynthesis	R_gltBD	13
Glutamine degradation	R_gln_deg	13
Glutamine biosynthesis	R_glnA	13.05
Acetyl-CoA synthetase	R_acs	13.06
Inorganic pyrophosphatase	R_ppa	13.08
Adenylate kinase	R_adk_atp	13.36
PEP carboxylase	R_ppc	14.22
Phosphofructokinase	R_pfk	14.9
Pyruvate kinase	R_pyk	15.89
Transhydrogenase	R_pnt2	22.45
Transhydrogenase	R_pnt1	25.5

Continued on next page

Enzyme/Pathway	Reaction	Uncertainty
Aspartate degradation	R_asp_deg	25.5
Aspartate biosynthesis	R_aspC	25.83
Asparagine biosynthesis	R_asnA	247.53
Asparagine degradation	R_asn_deg	247.53

Table 5.3: Flux uncertainty calculated using flux variability analysis for the base synthetic dataset during the second production phase (1 h to 3 h), normalized to the glucose consumption rate.

Effector	Effect	Target
Source		
Step 6 of AMP synthesis	R_A_syn_6	<0.01
Orotate synthase 1	R_or_syn_1	<0.01
Orotate synthase 2	R_or_syn_2	<0.01
Aldhyde-alcohol dehydrogenase	R_adhE_net	<0.01
RNA polymerase	Translation	<0.01
tRNA charging of phenylalanine	tRNA charging (PHE)	<0.01
tRNA charging of alanine	tRNA charging (ALA)	<0.01
tRNA charging of glutamine	tRNA charging (GLN)	<0.01
tRNA charging of threonine	tRNA charging (THR)	<0.01
tRNA charging of aspartate	tRNA charging (ASP)	<0.01
tRNA charging of glutamate	tRNA charging (GLU)	<0.01
tRNA charging of histidine	tRNA charging (HIS)	<0.01
tRNA charging of lysine	tRNA charging (LYS)	<0.01
tRNA charging of tyrosine	tRNA charging (TYR)	<0.01
tRNA charging of asparagine	tRNA charging (ASN)	<0.01
tRNA charging of serine	tRNA charging (SER)	<0.01
tRNA charging of methionine	tRNA charging (MET)	<0.01
tRNA charging of isoleucine	tRNA charging (ILE)	<0.01
tRNA charging of valine	tRNA charging (VAL)	<0.01
tRNA charging of proline	tRNA charging (PRO)	<0.01
tRNA charging of leucine	tRNA charging (LEU)	<0.01
tRNA charging of arginine	tRNA charging (ARG)	<0.01
tRNA charging of tryptophan	tRNA charging (TRP)	<0.01
tRNA charging of cysteine	tRNA charging (CYS)	<0.01
tRNA charging of glycine	tRNA charging (GLY)	<0.01
RNA polymerase	Translation initiation	<0.01
mRNA degradation	mRNA degradation	<0.01

Continued on next page

Enzyme	Reaction	Uncertainty
Glycine cleavage system	R_gly_fol_net	0.01
Transaldolase	R_talAB_net	0.02
Transketolase 1	R_tkt1_net	0.02
Ribulose-phosphate 3-epimerase	R_rpe_net	0.02
Ribose-5-phosphate isomerase	R_rpi_net	0.03
Transketolase 2	R_tkt2_net	0.04
Valine biosynthesis	R_val	0.05
Leucine biosynthesis	R_leu	0.05
Malate dehydrogenase	R_mdh_net	0.06
Triosephosphate isomerase	R_tpiA_net	0.07
Fructose 1,6-bisphosphate aldolase	R_fbaA_net	0.07
Glycine biosynthesis	R_gly_deg	0.09
Threonine degradation 2	R_thr_deg2	0.09
Phosphoglucose isomerase	R_pgi_net	0.09
Methylglyoxal degradation	R_mglx_deg	0.09
Methylenetetrahydrofolate dehydrogenase	R_mthfd_net	0.13
Glucose-6-phosphate dehydrogenase	R_zwf_net	0.18
Tyrosine biosynthesis	R_tyr	0.2
Enolase	R_eno_net	0.2
Phosphoglycerate mutase	R_gpm_net	0.2
Phosphoglycerate kinase	R_pgk_net	0.21
Glyceraldehyde-3-phosphate dehydrogenase	R_gapA_net	0.21
Methylenetetrahydrofolate dehydrogenase	R_mthfc_net	0.28
Metionine biosynthesis	R_met	0.34
Phosphate acetyltransferase	R_pta_net	0.4
Acetate kinase	R_ackA_net	0.41
Fumarate hydratase	R_fum_net	0.43
Phenylalanine biosynthesis	R_phe	0.53
Glucokinase	R_glk_atp	1
Step 5 of AMP synthesis	R_A_syn_5	1
Step 7 of AMP synthesis	R_A_syn_7	1

Continued on next page

Enzyme	Reaction	Uncertainty
OMP synthesis	R_omp_syn	1.09
Isoleucine biosynthesis	R_ile	1.13
Alanine biosynthesis	R_alaAC_net	1.49
Step 8 of AMP synthesis	R_A_syn_8	1.76
Methylenetetrahydrofolate reductase	R_mthfr2a	1.85
Isocitrate dehydrogenase	R_icd_net	1.87
Histidine biosynthesis	R_his	1.95
Step 4 of AMP synthesis	R_A_syn_4	2.02
Threonine degradation 3	R_thr_deg3	2.08
CTP synthetase 1	R_ctp_1	2.09
CTP synthetase 2	R_ctp_2	2.09
Lysine biosynthesis	R_lys	2.13
Lysine degradation	R_lys_deg	2.13
Lactate dehydrogenase	R_ldh_net	2.27
Tryptophan degradation	R_trp_deg	2.55
Tryptophan biosynthesis	R_trp	2.7
Aconitase	R_acn_net	2.82
UMP synthesis	R_ump_syn	2.85
Step 3 of AMP synthesis	R_A_syn_3	3.07
XMP synthase	R_xmp_syn	4.04
GMP synthetase	R_gmp_syn	4.04
Step 12 of AMP synthesis	R_A_syn_12	4.13
Step e of folate synthesis	R_fol_e	4.37
Step 2 of AMP synthesis	R_A_syn_2	4.74
Step 9 of AMP synthesis	R_A_syn_9	4.83
Step 10 of AMP synthesis	R_A_syn_10	4.83
Step 1 of AMP synthesis	R_A_syn_1	5.02
Glutamate dehydrogenase	R_gdhA_net	5.14
Phosphoribosylpyrophosphate synthase	R_prpp_syn	6.11
Step 4 of folate synthesis	R_fol_4	6.24
Step 3 of folate synthesis	R_fol_3	6.24

Continued on next page

Enzyme	Reaction	Uncertainty
Step 2b of folate synthesis	R_fol_2b	7.03
Step 2a of folate synthesis	R_fol_2a	8.04
Step 1 of folate synthesis	R_fol_1	9.04
Glycine biosynthesis	R_glyA	9.58
Chorismate synthesis	R_chor	13.98
Arginine degradation	R_arg_deg	14.97
Succinate thiokinase	R_sucCD	17.94
GABA degradation 1	R_gaba_deg1	17.95
GABA degradation 2	R_gaba_deg2	17.95
Glutamate degradation	R_glu_deg	17.95
2-Ketoglutarate dehydrogenase	R_sucAB	18.06
Arginine biosynthesis	R_arg	18.14
Threonine degradation 1	R_thr_deg1	19.06
Pyruvate formate lyase	R_pflAB	19.97
Threonine biosynthesis	R_thr	22.23
Malate synthase	R_aceB	28.81
Isocitrate lyase	R_aceA	28.81
Pyruvate dehydrogenase	R_pdh	29.86
6-Phosphogluconate dehydrogenase	R_gnd	31.9
Citrate synthase	R_gltA	32.2
Cysteine biosynthesis	R_cysEMK	48.93
Cysteine degradation	R_cys_deg	49.31
Proline biosynthesis	R_pro	61.2
2-Keto-3-deoxy-6-phospho-gluconate aldolase	R_eda	61.86
6-Phosphogluconate dehydrase	R_edd	61.86
Proline degradation	R_pro_deg	62.62
Serine degradation	R_ser_deg	68.5
6-Phosphogluconolactonase	R_pgl	70.12
Serine biosynthesis	R_serABC	72.05
Nucleotide diphosphatase (ATP)	R_atp_amp	74.22
Nucleotide diphosphatase (UTP)	R_utp_ump	74.22

Continued on next page

Enzyme	Reaction	Uncertainty
Nucleotide diphosphatase (GTP)	R_gtp_gmp	74.22
Nucleotide diphosphatase (CTP)	R_ctp_cmp	74.22
Cytidylate kinase	R_atp_cmp	74.74
Guanylate kinase	R_atp_gmp	76.11
UMP kinase	R_atp_ump	76.78
NADH dehydrogenase 1	R_ndh1	86.63
NADH:ubiquinone oxidoreductase	R_nuo	86.63
Malic enzyme A	R_maeA	86.77
NADH dehydrogenase 2	R_ndh2	87.01
Fumurate reductase	R_frd	87.01
Malic enzyme B	R_maeB	88.17
Succinate dehydrogenase	R_sdh	90.66
Cytochrome oxidase bo	R_cyo	92.08
Cytochrome oxidase bd	R_cyd	92.08
ATP synthase	R_atp	134.61
PEP synthase	R_pps	148.45
Asparagine biosynthesis	R_asnB	148.45
Glutamine degradation	R_gln_deg	148.45
Glutamate biosynthesis	R_gltBD	148.45
Acetyl-CoA synthetase	R_acs	148.45
Adenosinetriphosphatase	R_atp_adp	148.45
Fructose-1,6-bisphosphate aldolase	R_fdp	148.45
PEP carboxykinase	R_pck	148.45
Glutamine biosynthesis	R_glnA	149.67
Inorganic pyrophosphatase	R_ppa	150.79
Adenylate kinase	R_adk_atp	152.88
PEP carboxylase	R_ppc	158.68
Phosphofructokinase	R_pfk	161.64
Pyruvate kinase	R_pyk	170.95
Transhydrogenase	R_pnt2	258.03
Aspartate degradation	R_asp_deg	286.46

Continued on next page

Enzyme	Reaction	Uncertainty
Transhydrogenase	R_pnt1	286.46
Aspartate biosynthesis	R_aspC	290.86
Asparagine biosynthesis	R_asnA	4421.48
Asparagine degradation	R_asn_deg	4421.48

Table 5.4: Metabolites by frequency of appearance in the simulated annealing constraint sets.

Metabolite	Symbol	Frequency
alpha-D-Glucose	GLC	89.9%
Glucose 6-phosphate	G6P	89.9%
Citrate	CIT	85.7%
Isocitrate	ICIT	84.9%
Fumarate	FUM	84.5%
Fructose 6-phosphate	F6P	83.6%
6-Phospho-D-glucono-1,5-lactone	6PGL	81.5%
sedo-Heptulose 7-phosphate	S7P	79.4%
Alanine	ALA	77.7%
Guanosine triphosphate	GTP	77.3%
Malate	MAL	75.6%
D-Ribulose 5-phosphate	RU5P	74.8%
Erythrose 4-phosphate	E4P	73.1%
Adenosine diphosphate	ADP	72.3%
alpha-Ketoglutarate	AKG	71.8%
Uridine diphosphate	UDP	70.2%
Succinate	SUCC	69.7%
Cytidine monophosphate	CMP	69.3%
Guanosine diphosphate	GDP	67.2%
6-Phospho-D-gluconate	6PGC	67.2%
Arginine	ARG	66.8%
Ribose 5-phosphate	R5P	66.4%
Methionine	MET	65.5%
Glyoxylate	GLX	65.5%
Glutamine	GLN	63.9%
Phenylalanine	PHE	63.4%
Valine	VAL	62.6%

Table 5.5: Metabolites by frequency of appearance in the 57 best simulated annealing constraint sets, those with error at least three orders of magnitude lower than the base synthetic dataset.

Metabolite	Symbol	Frequency
D-Xylulose 5-phosphate	XU5P	100%
sedo-Heptulose 7-phosphate	S7P	100%
D-Ribulose 5-phosphate	RU5P	100%
Ribose 5-phosphate	R5P	100%
Oxaloacetate	OAA	100%
Isocitrate	ICIT	100%
alpha-D-Glucose	GLC	100%
Glucose 6-phosphate	G6P	100%
Glyceraldehyde 3-phosphate	G3P	100%
Fumarate	FUM	100%
Fructose 1,6-diphosphate	FDP	100%
Fructose 6-phosphate	F6P	100%
Erythrose 4-phosphate	E4P	100%
Dihydroxyacetone phosphate	DHAP	100%
Citrate	CIT	100%
Alanine	ALA	100%
6-Phospho-D-glucono-1,5-lactone	6PGL	100%
6-Phospho-D-gluconate	6PGC	100%
2-Dehydro-3-deoxy-D-gluconate 6-phosphate	2DDG6P	100%
Uridine triphosphate	UTP	98.2%
alpha-Ketoglutarate	AKG	98.2%
Succinate	SUCC	94.7%
Arginine	ARG	94.7%
3-Phosphoglycerate	3PG	94.7%
Guanosine triphosphate	GTP	91.2%
Uridine monophosphate	UMP	89.5%

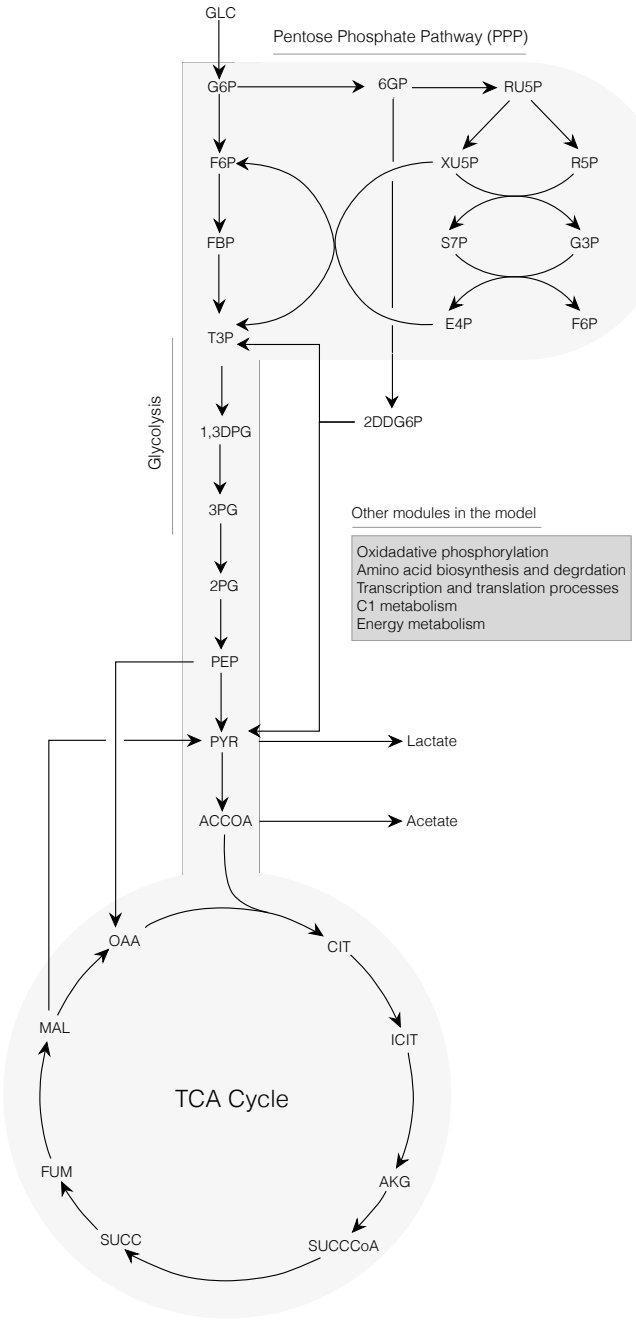

Figure 5.1: Schematic of the core portion of the cell-free *E. coli* metabolic network. The network consisted of 264 reactions and 146 metabolites. Metabolites of glycolysis, pentose phosphate pathway, Entner-Doudoroff pathway, and TCA cycle are shown. Metabolites of oxidative phosphorylation, amino acid biosynthesis and degradation, transcription/translation, chorismate metabolism, and energy metabolism are not shown.

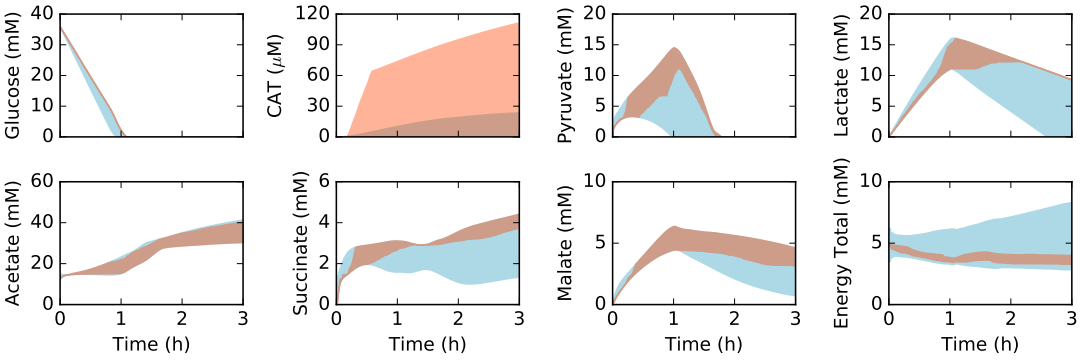

Figure 5.2: Simulated metabolite concentration versus synthetic data as a function of time. Central carbon metabolism, including glucose (substrate), CAT (product), and intermediates, as well as total concentration of energy species (energy total). The energy total denotes the summation of all energy species in the model (all bases and all phosphate states). The 95% confidence interval for the simulation conducted over the ensemble of transcription/translation parameter sets is shown in the orange shaded region, while the 95% confidence interval for the synthetic constraint data is shown in the blue shaded region. The synthetic data constraints were generated from the kinetic model of Horvath et al, which was trained using experimental measurements of the system simulated in this study [103].

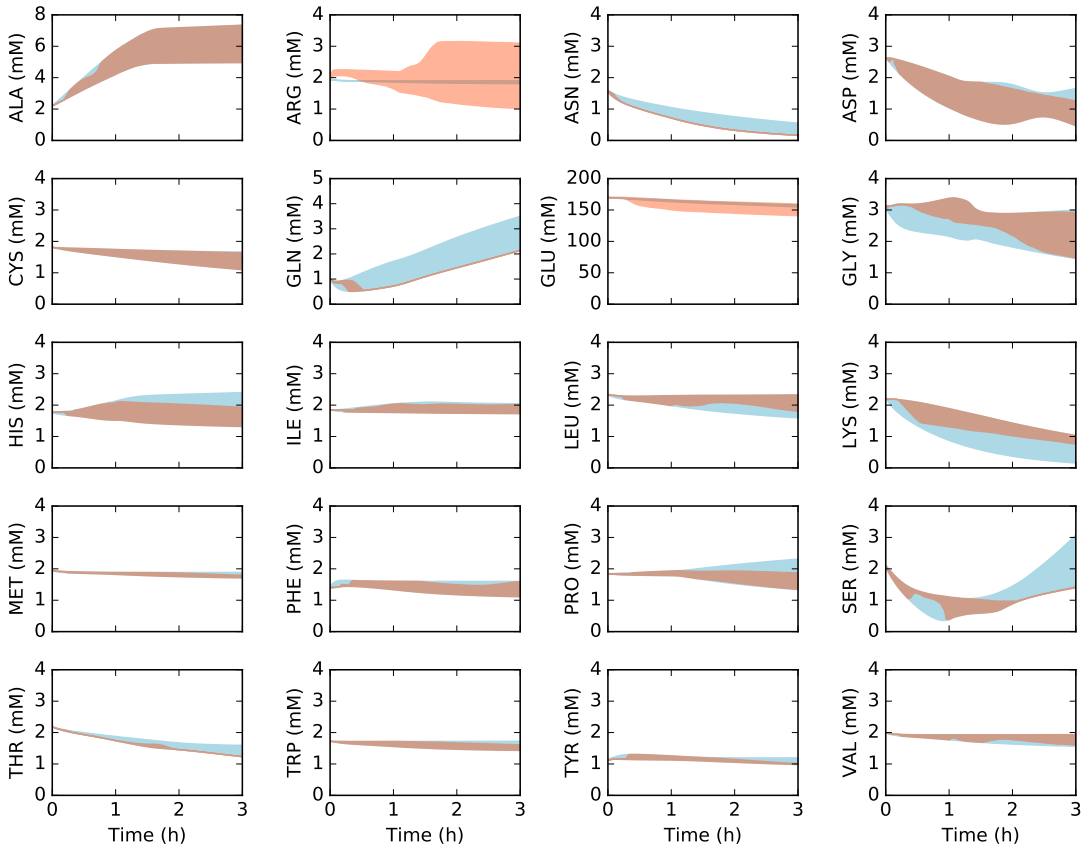

Figure 5.3: Simulation of amino acid concentration versus synthetic data as a function of time. The 95% confidence interval for the simulation conducted over the ensemble of transcription/translation parameter sets is shown in the orange shaded region, while the synthetic constraint data is shown in the blue shaded region. Arginine and glutamate were excluded from the constraint set. The synthetic data constraints were generated from the kinetic model of Horvath et al, which was trained using experimental measurements of the system simulated in this study [103].

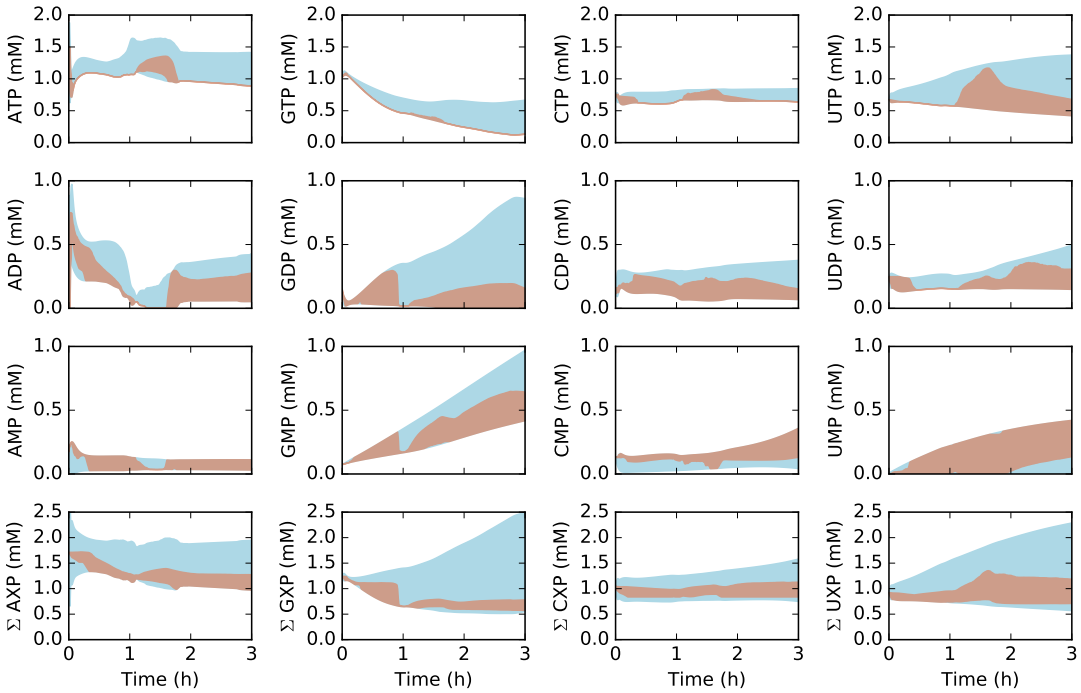

Figure 5.4: Simulation of energy species and energy totals by base versus synthetic data as a function of time. The 95% confidence interval for the simulation conducted over the ensemble of transcription/translation parameter sets is shown in the orange shaded region, while the synthetic constraint data is shown in the blue shaded region. The synthetic data constraints were generated from the kinetic model of Horvath et al, which was trained using experimental measurements of the system simulated in this study [103].

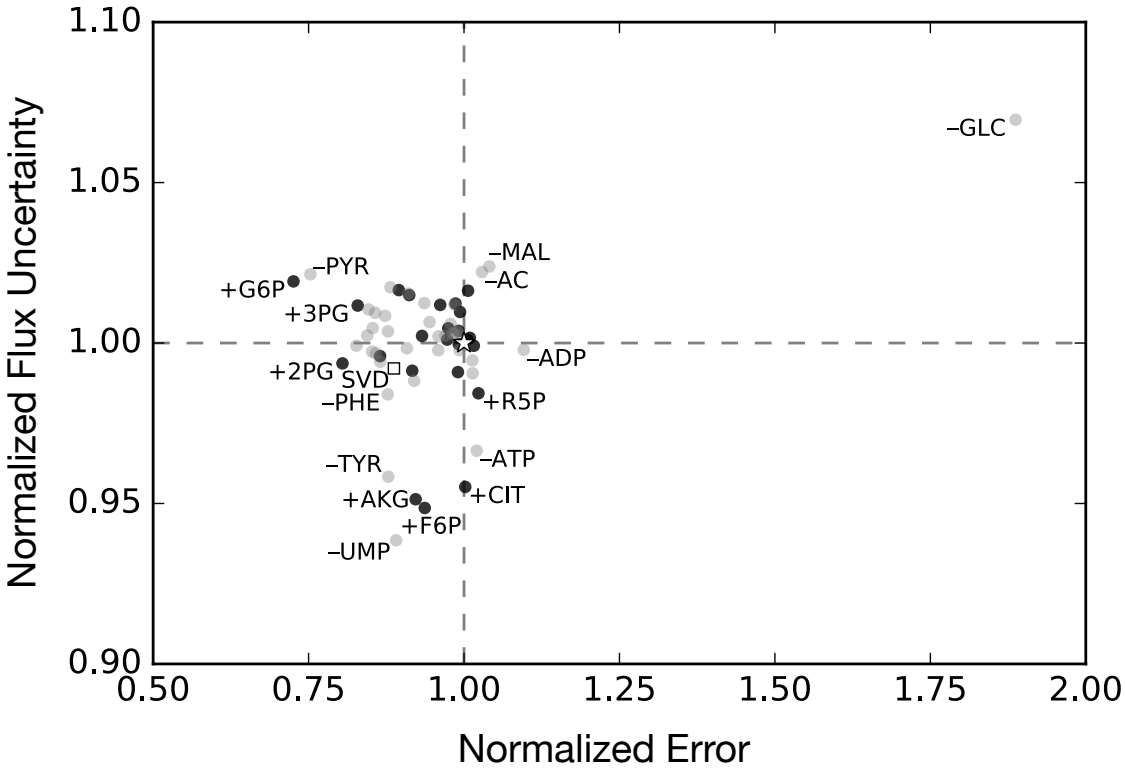

Figure 5.5: Flux uncertainty versus metabolite prediction error against synthetic data, normalized to the base case (white star), for exclusion (gray) and inclusion (black) metabolite constraint sets. The performance of the SVD-determined metabolite constraint set is shown by the white square.

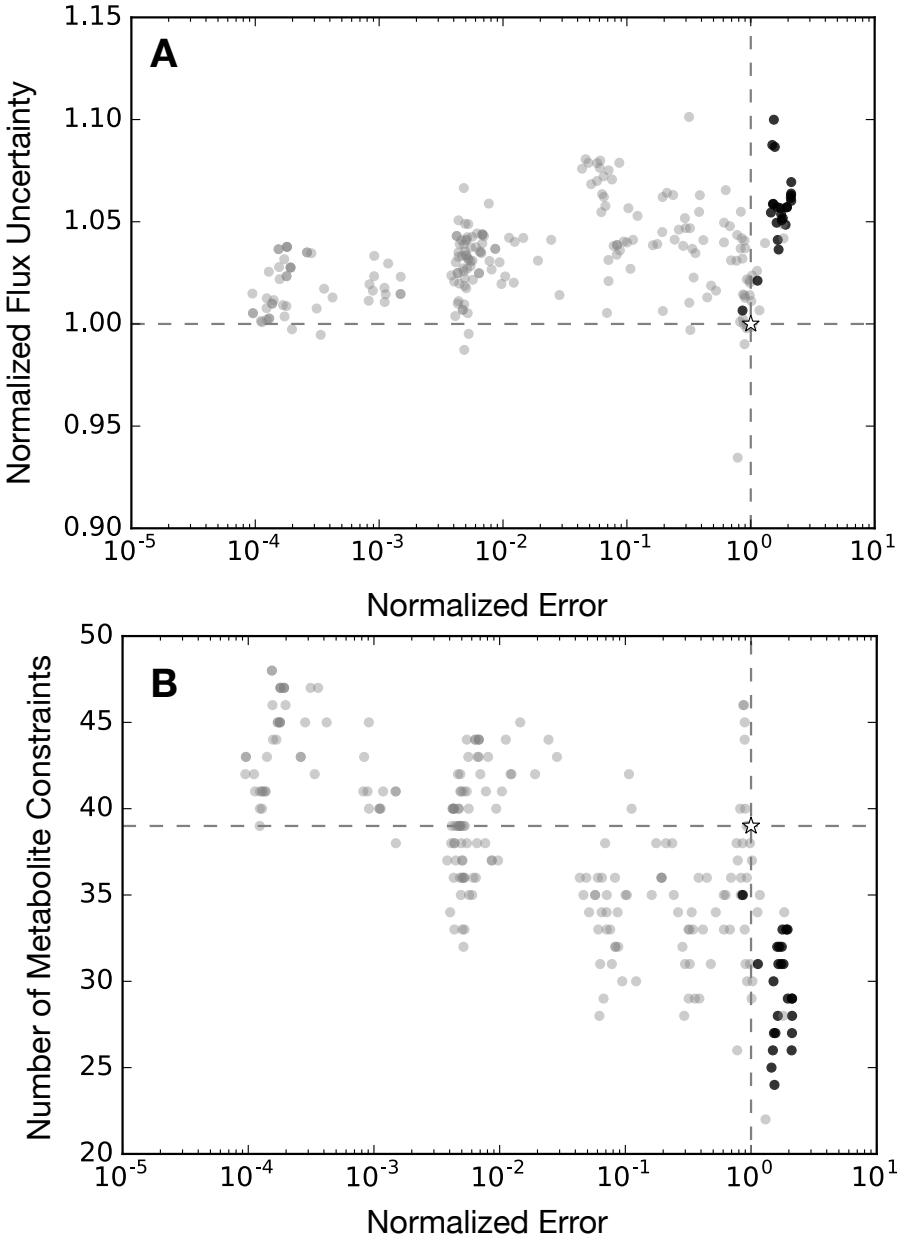

Figure 5.6: Flux uncertainty and metabolite prediction error for the simulated annealing experimental design approach. **A**: Normalized flux uncertainty versus normalized metabolite prediction error. **B**: Number of metabolite constraints versus normalized metabolite prediction error. Error was computed for the synthetic experimental designs normalized to the base synthetic dataset (white star). Sets that include glucose are show as gray circles, while those that do not are represented with black circles.

CHAPTER 6

METABOLIC CHARACTERIZATION OF ATRA-INDUCED HL60 DIFFERENTIATION

We developed liquid chromatography / mass spectrometry (LC/MS) methods for targeted metabolomics to characterize and quantify the changes in the metabolic profile caused by ATRA-induced differentiation. An ACQUITY UPLC BEH Amide Column was used on an ACQUITY UPLC H-Class System for compound separation at room temperature. The instrument used a single-quadrupole mass spectrometer (Waters, QDa) that was equipped with a pre-optimized electrospray ionization source.

To characterize metabolic profiles of ATRA-induced differentiation, we cultured ATRA-treated HL60 cells following standard protocol. Cells were counted and media samples were collected at 12 hour intervals for 72 hours. We characterized the glucose concentration using a Bayer Contour glucose meter, and lactate concentration using the LC/MS protocol. We observed that ATRA-treated cells lead to a decrease cell count, higher glucose concentration, and lower lactate concentration compared to the control (Fig. 6.1). To ensure that the cells were differentiated with ATRA treatment, flow cytometry was used to characterize the early myeloid differential marker, CD38 (Fig. 6.1). The specific growth, glucose consumption, and lactate production rates were determined to be: 0.37/hr, 0.125 mmol glucose/gDWhr, and 0.325 mmol lactate/gDWhr for untreated cultures, and 0.31/hr, 0.109 mmol glucose/gDWhr, and 0.502 mmol lactate/gDWhr for ATRA-treated cultures. The specific growth rate and glucose uptake rate were reduced with ATRA-induced differentiation as expected, however ATRA-treated cells produced more lactate compared to the control.

Amino acids were separated with a reverse-phase chromatography and quantified with a TUV detector at 260nm using a AccQ-Tag Ultra Derivatization Kit (Waters Corporation, Milford, MA) (Figs. 6.2-6.5).

To better understand the metabolic rewiring associated with differentiation, we used the specific rates above as extracellular constraints in a flux balance analysis (FBA) problem (Fig. 6.6). The metabolic reconstruction was adopted from Palsson and coworker's RECON 2.2 [219]. The network is a genome scale reconstruction with 6048 metabolites and 11567 reactions. The increased lactate production with ATRA-treated cells was unexpected due to the decreased glucose utilization and growth rate. Interestingly, to account for the increased lactate production, while decreasing glucose, the FBA predicted an increased glutamine utilization. This increased glutamine consumption provided a secondary substrate to replenish TCA cycle metabolites that are required for biomass synthesis and oxidative phosphorylation. We experienced challenges in determining the intracellular flux distribution due to the large number of possible solutions to the linear programming problem. To overcome this challenge, we began expanding our experimental efforts to incorporate nucleic acid quantification and enzyme activities to introduce more constraints into the model (Fig.6.7).

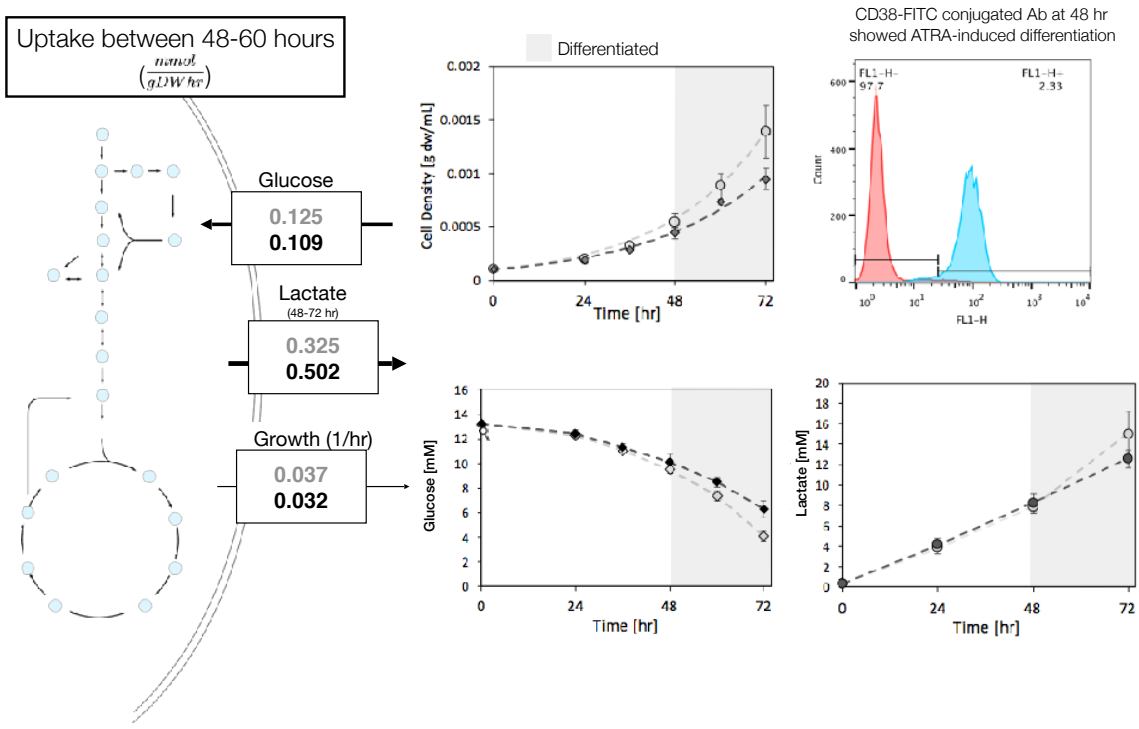

Figure 6.1: HL60 culture seeded at 100k cells/mL cultured for 72 hours. Growth curve, glucose and lactate media concentrations. The light gray represent the control while the dark gray are the ATRA-treated cells. The error bars represent one standard deviation of three biological replicates. The arrows represents the exchange of metabolic fluxes.

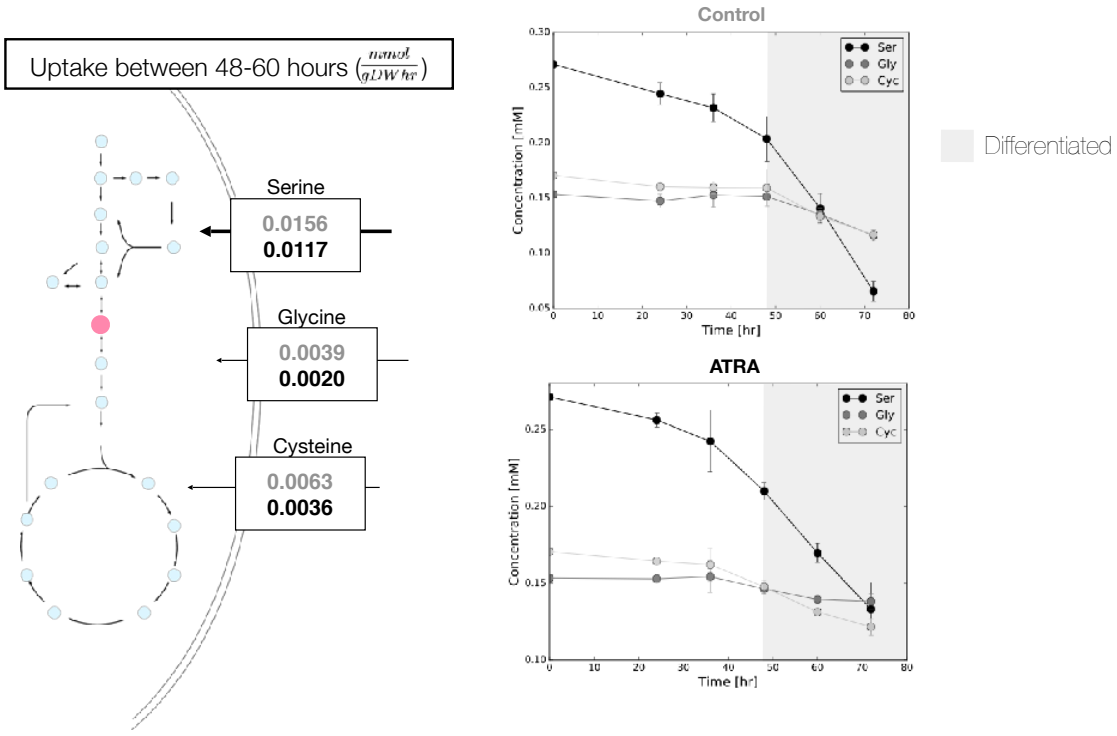

Figure 6.2: HL60 culture seeded at 100k cells/mL cultured for 72 hours. Time course of serine, glycine, and cysteine media concentrations. The light gray represent the control while the dark gray are the ATRA-treated cells. The error bars represent one standard deviation of three biological replicates. The arrows represents the exchange of metabolic fluxes.

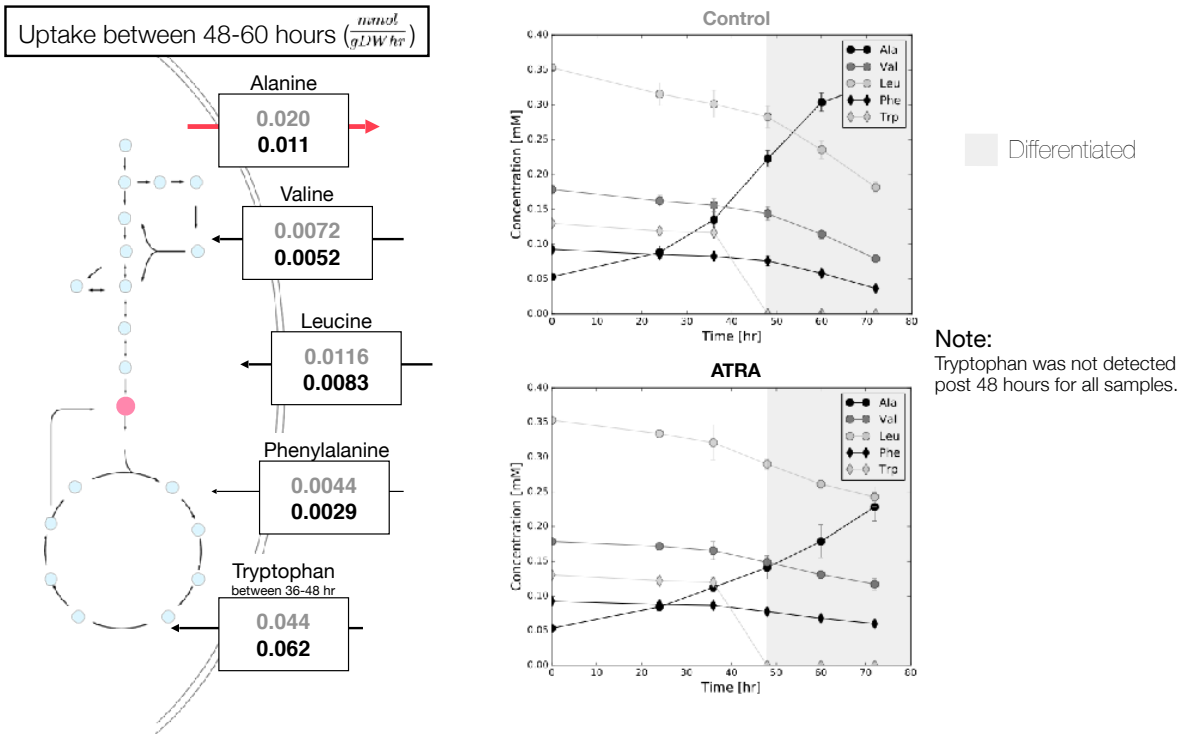

Figure 6.3: HL60 culture seeded at 100k cells/mL cultured for 72 hours. Time course of alanine, valine, leucine, phenylalanine, and tryptophan media concentrations. The light gray represent the control while the dark gray are the ATRA-treated cells. The error bars represent one standard deviation of three biological replicates. The arrows represents the exchange of metabolic fluxes.

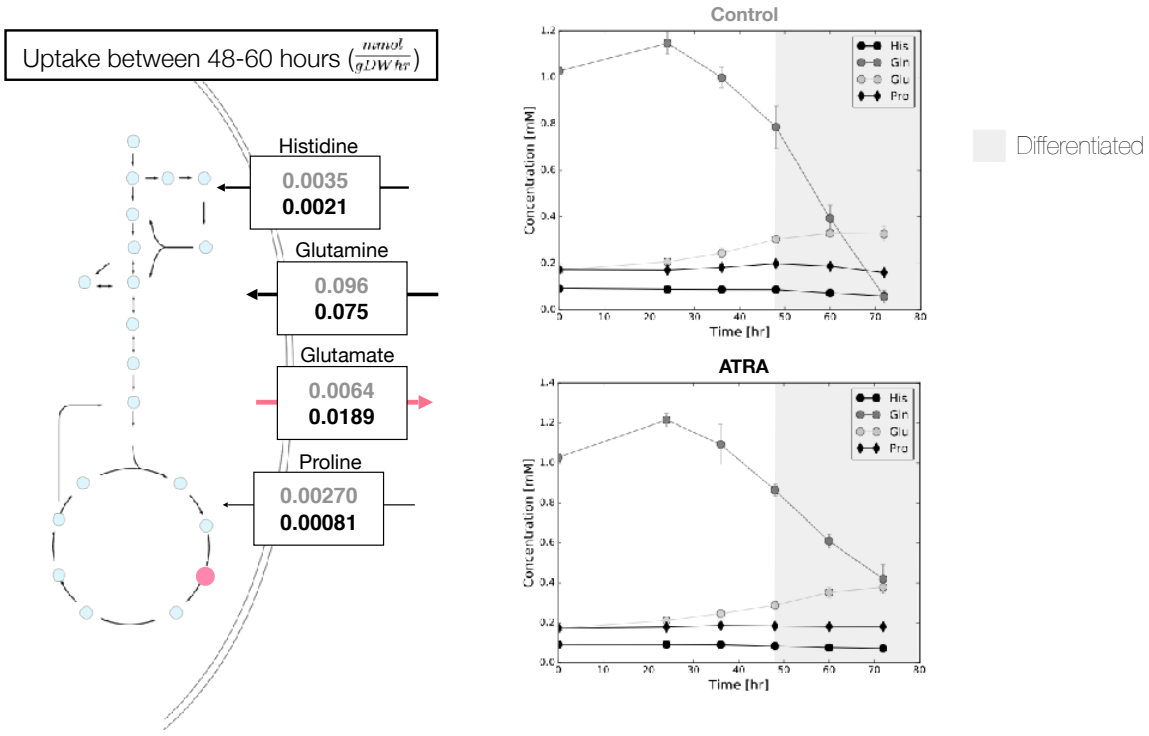

Figure 6.4: HL60 culture seeded at 100k cells/mL cultured for 72 hours. Time course of histidine, glutamine, glutamate, and proline media concentration. The light gray represent untreated cultures while the dark gray are the ATRA-treated cultures. The error bars represent one standard deviation of three biological replicates.

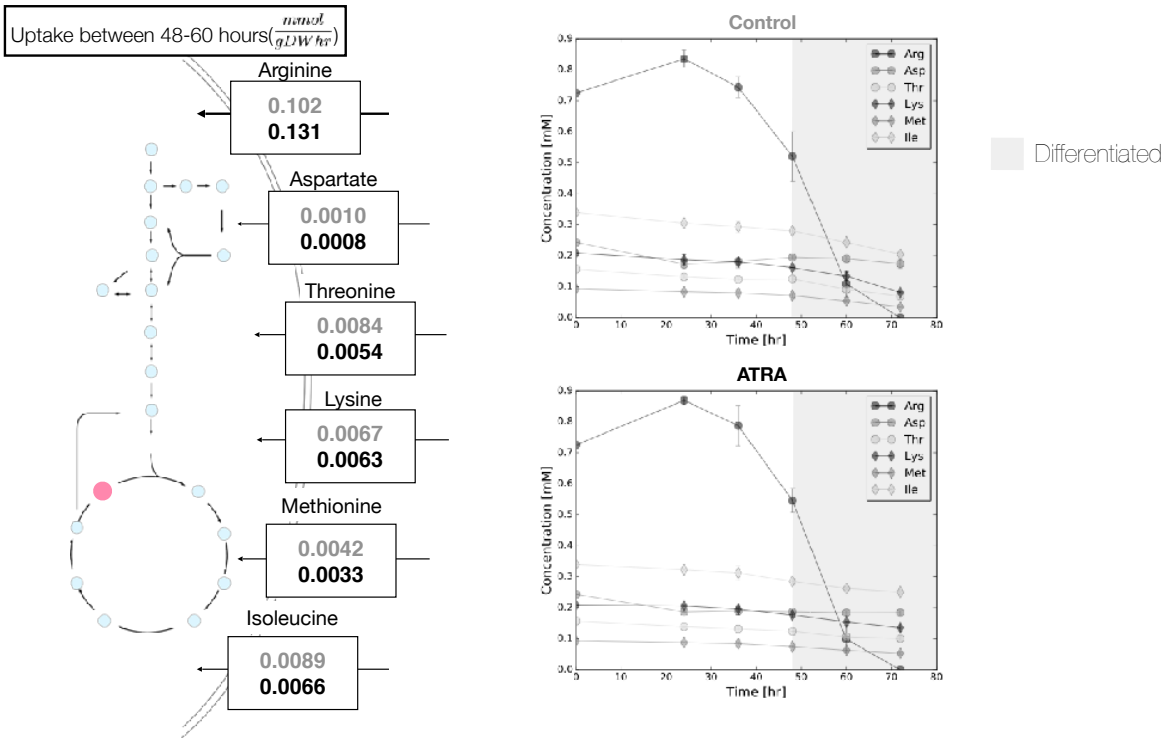

Figure 6.5: HL60 culture seeded at 100k cells/mL cultured for 72 hours. Time course of arginine, aspartate, threonine, lysine, methionine, isoleucine media concentration. The light gray represent untreated cultures while the dark gray are the ATRA-treated cultures. The error bars represent one standard deviation of three biological replicates.

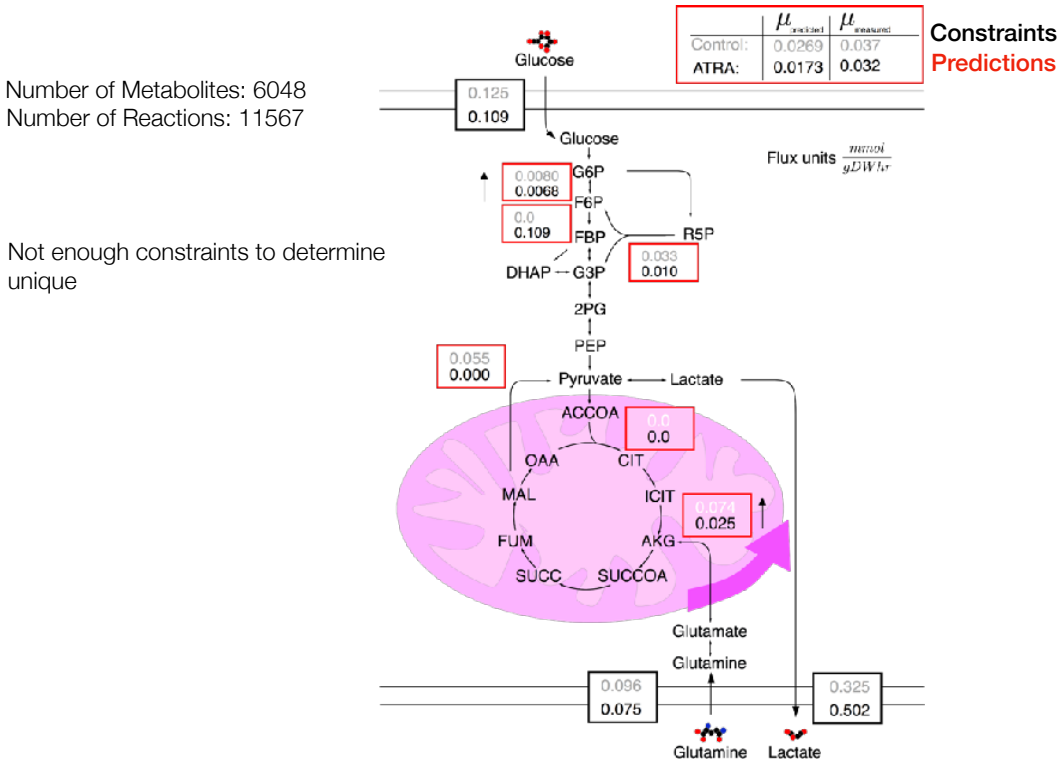

Figure 6.6: Flux balance analysis model using RECON 2.2 with constraints from extracellular media measurements. Light gray indicates untreated HL60 fluxes while Dark gray indicates ATRA-treated HL60 fluxes. All fluxes are in the units of mmol/gDW.

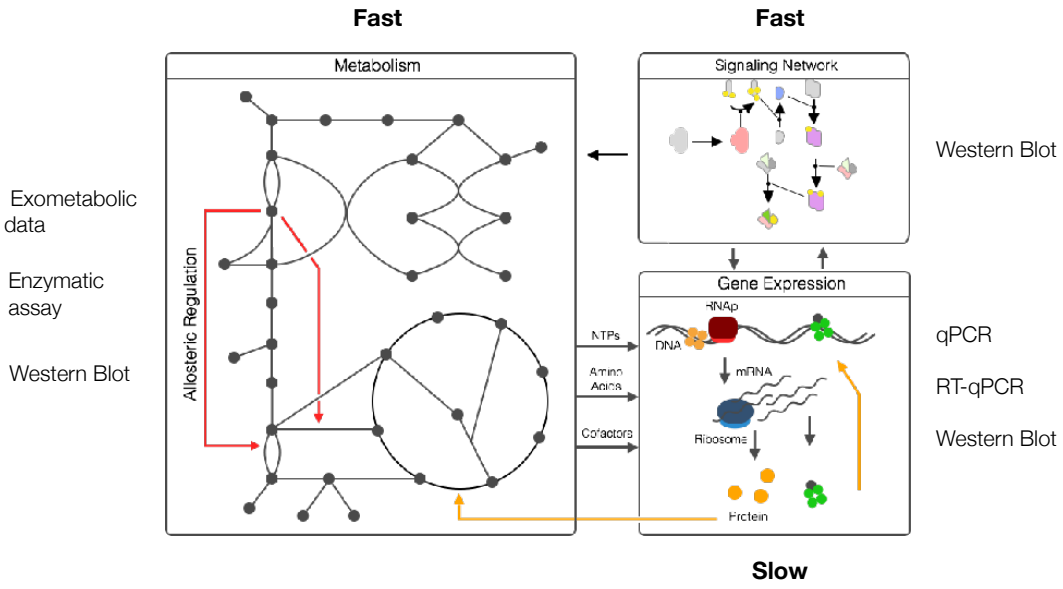

Figure 6.7: Schematic approach in connecting the different hierarchy of cellular decision making.

CHAPTER 7

ON GOING AND FUTURE WORKS

7.1 Response Resistance and Rescue - Metabolic characterization of ATRA-induced differentiation

The main motivation of this work was to create an integrated systems biology approach to better understand resistance in cancer and the how we can use combination therapy to alleviate make the cells responsive. In the future, we expand the metabolic characterization of ATRA-resistant cell lines [118]. An initial control would be conducted using Am580, a nonmetabolizable analog of ATRA to ensure ATRA-associated signaling is occurring throughout the time-course. The resistant cell lines are characterized by loss of RA-inducible G1/G0 arrest, CD11b expression, inducible oxidative metabolism and p47(phox) expression. In a previous study, Jensen and coworkers showed that ATRA-treated RA-resistant HL60 exhibits sustained MEK/ERK activation, and one of the two sequentially emergent resistant lines retains ATRA-inducible CD38 expression. Other signaling events that define the wild-type (WT) response are also compromised, including c-Raf phosphorylation [118]. Jensen and coworkers also showed that SFK inhibitors such as PP2 treatment combined with ATRA initiated characteristics of wild type HL60s but were potentially incapable of developing inducible oxidative metabolism [118].

We plan to develop methods to quantify glutamine using the LC/MS and fully characterize conditioned media samples for a more comprehensive understanding of nutrient utilization of HL60 metabolism. With the additional

metabolic measurements in the conditioned media, we would expand our core metabolic model to incorporate amino acid synthesis and uptake to provide a more accurate flux distribution. In addition, other metabolic targets such as hexokinase II (total and active Thr473) and lactate dehydrogenase will be probed as well, since they represent enzymes that catalyze the first and last reaction of aerobic glycolysis, respectively. To incorporate new constraints and nutrient uptake into the flux balance analysis, the core network needs to be expanded to incorporate additional exchange biosynthetic reactions. One way of achieving this is to incorporate new pathways manually. Another approach is to adopt the NetworkReducer of Erdich and coworkers, which can be used to systematically reduce a genome-scale model into a more manageable core network with user specifications [64]. To validate the FBA model, we propose probing for metabolic enzymes that (1) showed zero flux, and (2) had large change in flux between treatments of ATRA and/or SFK inhibitor treatment. Zero-flux enzymes include: phosphofructose kinase (PFK) that converts G6P to F6P in upper glycolysis, citrate synthase, aconitase 2, and isocitrate dehydrogenase (IDH) 2/3 that catalyzes the first 3 steps in the TCA cycle within the mitochondria. In addition, we propose conducting splitting analysis using enzymatic assay kits of the respective enzymes at a branch point. To better understand the mechanisms of the metabolic reprogramming associated with ATRA-induced differentiation or lack thereof, we aim to probe for specific signaling molecules associated with the aerobic glycolysis metabolism, especially PI3K/AKT/mTOR pathway and transcription factor downstream of MAPK such as MYC and HIF1-α [65, 145, 65, 170, 203]. We also aim to characterize the rate of transcription through quantitative real time polymerase chain reaction (RT-qPCR) analysis. This method will offer an additional quantitative measurement that

can bridge the gap between measurements in signaling molecules, transcription factors, and enzyme expression. More importantly, RT-PCR will be crucial in determining alternative splicing variants, for instance pyruvate kinase isoform 2 (PKM2), which is known to induce LDH [159]. Ultimately, transcription data and mRNA levels provide an independent validation for enzyme expression, and training data for gene expression models.

Lastly, with the development of our experimental methods, we propose to conduct an set of exploratory experiments that look at enzyme translocation and noncanonical epigenic functions of metabolic enzymes in the nucleus. This can be done using the same experiments with the additional step of nuclear fractionation. One particular metabolic enzyme of interest is pyruvate kinase (specifically PKM2) and pyruvate dehydrogenase complex (PDC) that can be translocated to the nucleus and form complexes with histone acetyltransferases (HATs) and provide acetyl-coa (normally impermeable to mitochondria membrane and unstable) for histone acetylation and gene expression [218, 166, 16].

7.2 Connecting Signaling to Metabolism using Kinetic and Constraint-based Modeling

The Warburg effect, the metabolic phenomenon of increased glucose uptake that is fermented into lactate in the presence of oxygen, has been observed across all cancer cells and is seen as a hallmark of cancer metabolism. One hypothesis suggests that expression of TCA enzymes is too costly to keep up with the dilution attributed by the uncontrolled cell division. We propose an integrated platform that integrates signaling , gene expression, and metabolism . The novel

modeling platform aims to capture cellular decision making at a global level for a more comprehensive understanding of phenotypic changes such as differentiation or metabolic programming. The main purpose of this model is to implement regulatory layers on top of the flux balance analysis model to couple gene expression data such as enzyme and RNA levels to metabolic fluxes. Having additional layers of regulation and constraints would greatly reduce the solution space and increase the predictive power of the model.

For this comprehensive model to include signaling, gene expression, and metabolism, the individual components or modules must be constructed and then integrated together. For the signaling module, we propose adapting Lequieu and coworkers' signaling model of insulin induced translation initiation [144]. The model has 823 parameters (573 kinetic parameters and 250 initial conditions) and a comprehensive representation including phosphorylation, activation, and complex formation of key species in the P13K/AKT/mTOR pathway that lead to downstream formation of 80S initiation complex. Although this model was developed to study insulin response, the core components of the PI3K/AKT/mTOR signaling can be adopted and expanded to other downstream transcription factor targets such as MYC and HIF1-α.

For the expression of transcription factors and metabolic enzymes, we propose the adoption of the effective kinetic modeling framework used in gene regulatory networks. We propose a model of metabolism, transcription, and translation in an HL60 metabolic network, encoded as a system of ordinary dif-

ferential equations using the same formulations as previously done:

$$\frac{dx_i}{dt} = \sum_{j=1}^{\mathcal{R}} \sigma_{ij} r_j(\mathbf{x}, \epsilon, \mathbf{k}) - x_i \mu \qquad i = 1, 2, \ldots, \mathcal{M} \tag{7.1}$$

$$\frac{dm_j}{dt} = r_{T,j} - \left(\mu + \theta_{m,j}\right) m_j + \lambda_j \tag{7.2}$$

$$\frac{dp_j}{dt} = r_{X,j} - \left(\mu + \theta_{p,j}\right) p_j \tag{7.3}$$

$$\tag{7.4}$$

The quantities x_i, m_j, and p_j denote metabolite i, mRNA species j, and protein j, respectively. The quantity $\mathcal{M}$ denotes the number of metabolites. The quantity $r_j(\mathbf{x}, \epsilon, \mathbf{k})$ denotes the rate of reaction j. Typically, reaction j is a non-linear function of metabolite and enzyme abundance, as well as unknown kinetic parameters $\mathbf{k}$ ($\mathcal{K} \times 1$). The quantity σ_{ij} denotes the stoichiometric coefficient for species i in reaction j. If $\sigma_{ij} > 0$, metabolite i is produced by reaction j. Conversely, if $\sigma_{ij} < 0$, metabolite i is consumed by reaction j, while $\sigma_{ij} = 0$ indicates metabolite i is not connected with reaction j. The rate of reaction ,$r_j(\mathbf{x}, \epsilon, \mathbf{k})$, is a function of active enzyme level for the specific reaction j:

$$r_j = k_{cat} E_j^* \mathbf{x} \tag{7.5}$$

$$E_j^* = E_j \alpha_j \tag{7.6}$$

Although mass action kinetics is shown above, other forms of kinetics can also be implemented. The variable k_{cat} is a kinetic constant that can be determined through enzymatic assays for each enzyme of interest, or estimated based on literature values. E_j^* is the active form of E_j, and its level can be directly measured or determined as a function α_j of post-translational modulators.

We plan to solve the metabolic model equations using a integrated dynamic constraint-based approach in which the rates of the signaling events, and transcription and translation of metabolic enzymes and fluxes, are estimated by solving a convex optimization subproblem at each time step k. We will pose this subproblem as a linear programming problem for convenience; however, this does not need to be the case. A key advantage of proposing the subproblem as a linear programming problem is the lack of parameters and parameter estimation.

The convex subproblem will consist of an objective function, material balances, and flux bound constraints. The linear objective function:

$$\min_{r_1,\ldots,r_C} \left(\sum_{j=1}^{C} c_j r_j(k) \right) \tag{7.7}$$

is minimized subject to the material balance constraints:

$$\begin{aligned} \frac{\theta_i^{min}}{h} \leq \sum_{j=1}^{C} \sigma_{ij} r_j(k) \leq \frac{\theta_i^{max}}{h} \qquad i = 1, \ldots, \mathcal{M} \\ \theta_i^{min/max} \equiv x_{i,\,k+1}^{min/max} - x_{i,\,k} \equiv x_i^{min/max}(t+h) - x_i(t) \end{aligned} \tag{7.8}$$

and flux bounds constraints:

$$\mathcal{L}_j(\mathbf{x}) \leq r_j \leq \mathcal{U}_j(\mathbf{x}) \qquad j = 1, \ldots, C$$

at each time step k. The user configurable parameters govern the maximum rate of increase and decrease of the ith species derived from the maximum and minimum concentrations, $x_i^{min/max}$. The $\mathcal{L}_j(\mathbf{x})$ and $\mathcal{U}_j(\mathbf{x})$ functions encode the

lower and upper bound for the jth flux, while h denotes the time step size. In this subproblem, the intracellular metabolite level does not need to be a steady state, allowing the model to predict accumulation and depletion of metabolites that may be caused by signaling and a shift in gene expression.

7.3 Expansion of Framework into Triple Negative Breast Cancer

The homeobox domain transcription factor NANOG, a key regulator of embryonic development and cellular reprogramming, is broadly expressed in human cancers. In addition to promoting self-renewal and long-term proliferative potential of stem-like cancer cells, NANOG-mediated oncogenic reprogramming may underlie clinical manifestations of malignant disease.In this study, we characterized MDA-MB-231 cells at three different NANOG expressions: Null, Low, and High. We quantified the extracellular metabolite concentration within the media, and used the results are extracellular-exchange constraints for our constraint-based modeling to interrogate the intracellular flux distributions.

A reporter line of MDA-MB-231 cells developed by Thiagarajan and coworkers were used for this study [224]. These cells were transfected with NANOG-GFP gene and the population were sorted into Null, Low, and High referring to: non GFP expression, bottom and top 5% GFP-positive of the population. Each sample was then cultured for 72 hours with media samples collected and analyzed. The media from the time-course samples was analyzed for glucose, lactate, and 20 amino acids using LC/MS analytical techniques. The rate of change was estimated using time course data of the cell count and metabolite

concentrations. The extracellular constraints for the FBA model were generated by sampling the statistics of the rate of change at the time of interest and setting the upper and lower limits to a standard deviation away from the sampled average.

The cell mass, glucose, and lactate measurements were measured (Fig. 7.1). With increasing NANOG expression, our analysis showed the following pathways increased per glucose consumed: upper and lower glycolysis, pyruvate transport into the TCA cycle, and malate to OAA. Similarly, we observed a decreasing pattern with NANOG expression in: glutamine uptake, oxygen consumption, electron transport chain, magic enzyme, and shunt from G6P into pentose phosphate and G1P.

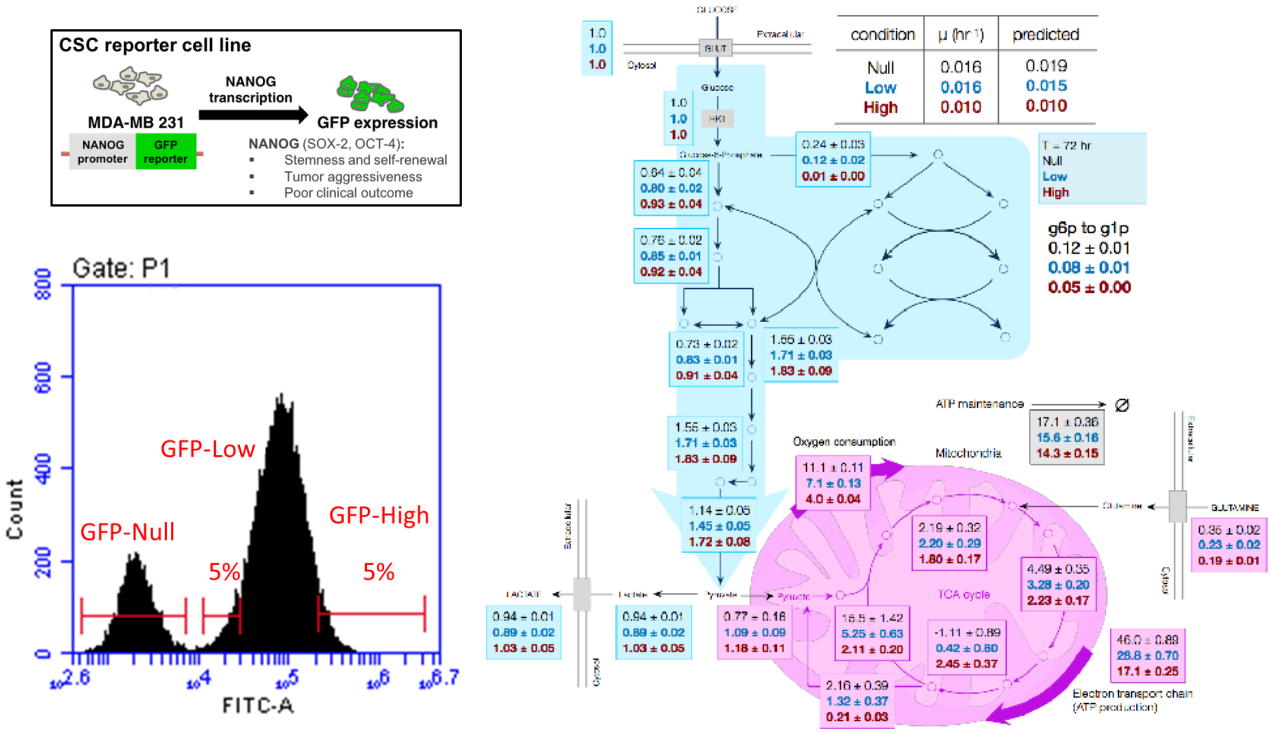

Figure 7.1: The MDA-MB-231 reporter cell line consists of a transfected vector that contains a NANOG promotor directly upstream of GFP. Therefore in the presence of NANOG, green fluorescent protein (GFP) will be produced, differentiation the high expressing NANOG cells from the low and null in the total population. In this study, we defined the high and low NANOG-expressing cells to be top and bottom 5% of the GFP positive cells, and null to be the subpopulation that are GFP negative. Leveraging our metabolomic efforts, we used the change in metabolite concentrations in the media as constraints for the FBA model to determine intracellular flux estimates for each subpopulations.

CHAPTER 8

CONCLUDING REMARKS

Early on in my graduate studies, I realize what fascinated me was how cells made decisions, adapting to their environment. The way they process information can lead to variety of changes from metabolic reprogramming to differentiation. This began my pursuit to understand how cells are regulated from signal transduction, to gene expression, post-translation modification, and metabolism. Leveraging the existing knowledge we had in signal transduction and metabolism, my PhD focused on addressing how environmental cues trigger signaling transduction and how those signals influences downstream metabolic programming in the context of metabolic diseases, specifically cancer. Having a mechanistic understanding between signaling and metabolism will further our understanding of these diseases. More importantly, computational models that incorporate mechanistic understanding of multiple layers of cellular decision making can be used to design combination therapies which combine multiple treatment modalities.

APPENDIX A

MODELS MADE AVAILABLE

Model	URL
HL60 ATRA Transcriptional Factor Model	https://github.com/varnerlab/cRAF-reduced-model-manuscript
HL60 ATRA&D3 Transcriptional Factor Model	https://github.com/varnerlab/HL60-TF-Model_v4
Cancer Metabolism Model	https://github.com/varnerlab/Core-Cancer-Model
Cell Free Protein Synthesis Model	https://github.com/varnerlab/publication_cell_free_dFBA_repository
Human Complement Model	https://github.com/varnerlab/Complement_model_repository

BIBLIOGRAPHY

[1] Targeted cancer therapies.

[2] Richat Abbas and Poe-Hirr Hsyu. Clinical pharmacokinetics and pharmacodynamics of bosutinib. *Clinical pharmacokinetics*, 55(10):1191–1204, 2016.

[3] Timothy E Allen and Bernhard Ø Palsson. Sequence-based analysis of metabolic demands for protein synthesis in prokaryotes. *J Theor Biol*, 220(1):1–18, Jan 2003.

[4] S Altiok, M Xu, and B M Spiegelman. Ppargamma induces cell cycle withdrawal: inhibition of e2f/dp dna-binding activity via down-regulation of pp2a. *Genes Dev*, 11(15):1987–98, Aug 1997.

[5] E Z Bagci, Y Vodovotz, T R Billiar, G B Ermentrout, and I Bahar. Bistability in apoptosis: roles of bax, bcl-2, and mitochondrial permeability transition pores. *Biophys J*, 90:1546–59, 2006.

[6] James E Bailey. Mathematical modeling and analysis in biochemical engineering: past accomplishments and future opportunities. *Biotechnology progress*, 14(1):8–20, 1998.

[7] James E Balmer and Rune Blomhoff. Gene expression regulation by retinoic acid. *J Lipid Res*, 43(11):1773–808, Nov 2002.

[8] N Barkal and Stan Leibler. Robustness in simple biochemical networks. *Nature*, 387(6636):913–917, 1997.

[9] B Bauvois, L Durant, J Laboureau, E Barthélémy, D Rouillard, G Boulla, and P Deterre. Upregulation of cd38 gene expression in leukemic b cells by interferon types i and ii. *J Interferon Cytokine Res*, 19(9):1059–66, Sep 1999.

[10] Qasim K Beg, Alexei Vazquez, Jason Ernst, Marcio A de Menezes, Ziv Bar-Joseph, A-L Barabási, and Zoltán N Oltvai. Intracellular crowding defines the mode and sequence of substrate uptake by escherichia coli and constrains its metabolic activity. *Proceedings of the National Academy of Sciences*, 104(31):12663–12668, 2007.

[11] G Behre, A J Whitmarsh, M P Coghlan, T Hoang, C L Carpenter, D E Zhang, R J Davis, and D G Tenen. c-jun is a jnk-independent

coactivator of the pu.1 transcription factor. *J Biol Chem*, 274(8):4939–46, Feb 1999.

[12] Jeff Bezanson, Alan Edelman, Stefan Karpinski, and Viral B. Shah. Julia: A fresh approach to numerical computing. *CoRR*, abs/1411.1607, 2014.

[13] Lacramioara Bintu, Nicolas E Buchler, Hernan G Garcia, Ulrich Gerland, Terrence Hwa, Jane Kondev, and Rob Phillips. Transcriptional regulation by the numbers: models. *Current Opinion in Genetics & Development*, 15(2):116–24, 2005.

[14] Karl Walter Bock. 2,3,7,8-tetrachlorodibenzo-p-dioxin (tcdd)-mediated deregulation of myeloid and sebaceous gland stem/progenitor cell homeostasis. *Arch Toxicol*, 91(6):2295–2301, Jun 2017.

[15] K Boström, M Wettesten, J Borén, G Bondjers, O Wiklund, and S O Olofsson. Pulse-chase studies of the synthesis and intracellular transport of apolipoprotein b-100 in hep g2 cells. *J Biol Chem*, 261(29):13800–6, Oct 1986.

[16] Aristeidis E Boukouris, Sotirios D Zervopoulos, and Evangelos D Michelakis. Metabolic enzymes moonlighting in the nucleus: metabolic regulation of gene transcription. *Trends in biochemical sciences*, 41(8):712–730, 2016.

[17] Derek P Brazil and Brian A Hemmings. Ten years of protein kinase b signalling: a hard akt to follow. *Trends in biochemical sciences*, 26(11):657–664, 2001.

[18] T. R. Breitman, S. E. Selonick, and S. J. Collins. Induction of differentiation of the human promyelocytic leukemia cell line (HL-60) by retinoic acid. *Proc Natl Acad Sci U S A*, 77:2936–2940, 1980.

[19] S. C. Brooks, S. Kazmer, A. A. Levin, and A. Yen. Myeloid differentiation and retinoblastoma phosphorylation changes in HL-60 cells induced by retinoic acid receptor- and retinoid X receptor-selective retinoic acid analogs. *Blood*, 87:227–237, 1996.

[20] Dennis Bruemmer, Fen Yin, Joey Liu, Joel P Berger, Toshiyuki Sakai, Florian Blaschke, Eckart Fleck, Andre J Van Herle, Barry M Forman,

and Ronald E Law. Regulation of the growth arrest and dna damage-inducible gene 45 (gadd45) by peroxisome proliferator-activated receptor gamma in vascular smooth muscle cells. *Circ Res*, 93(4):e38–47, Aug 2003.

[21] Rodica P Bunaciu, Holly A Jensen, Robert J MacDonald, Dorian H LaTocha, Jeffrey D Varner, and Andrew Yen. 6-formylindolo(3,2-b)carbazole (ficz) modulates the signalsome responsible for ra-induced differentiation of hl-60 myeloblastic leukemia cells. *PLoS One*, 10(8):e0135668, 2015.

[22] Rodica P Bunaciu and Andrew Yen. Activation of the aryl hydrocarbon receptor ahr promotes retinoic acid-induced differentiation of myeloblastic leukemia cells by restricting expression of the stem cell transcription factor oct4. *Cancer Res*, 71(6):2371–80, Mar 2011.

[23] Rodica P Bunaciu and Andrew Yen. 6-formylindolo (3,2-b)carbazole (ficz) enhances retinoic acid (ra)-induced differentiation of hl-60 myeloblastic leukemia cells. *Mol Cancer*, 12:39, 2013.

[24] Anthony P Burgard and Costas D Maranas. Optimization-based framework for inferring and testing hypothesized metabolic objective functions. *Biotechnology and bioengineering*, 82(6):670–677, 2003.

[25] Nathan Bushue and Yu-Jui Yvonne Wan. Retinoid pathway and cancer therapeutics. *Adv Drug Deliv Rev*, 62(13):1285–98, Oct 2010.

[26] Mariajose Castellanos, David B Wilson, and Michael L Shuler. A modular minimal cell model: purine and pyrimidine transport and metabolism. *Proceedings of the National Academy of Sciences of the United States of America*, 101(17):6681–6686, 2004.

[27] Fei Chen, Qing Wang, Xuening Wang, and George P Studzinski. Up-regulation of egr1 by 1,25-dihydroxyvitamin d3 contributes to increased expression of p35 activator of cyclin-dependent kinase 5 and consequent onset of the terminal phase of hl60 cell differentiation. *Cancer Res*, 64(15):5425–33, Aug 2004.

[28] H Chen, D Ray-Gallet, P Zhang, C J Hetherington, D A Gonzalez, D E Zhang, F Moreau-Gachelin, and D G Tenen. Pu.1 (spi-1) autoregulates its expression in myeloid cells. *Oncogene*, 11(8):1549–60, Oct 1995.

[29] H Chen, P Zhang, H S Radomska, C J Hetherington, D E Zhang, and D G Tenen. Octamer binding factors and their coactivator can activate the murine pu.1 (spi-1) promoter. *J Biol Chem*, 271(26):15743–52, Jun 1996.

[30] Florence S G Cheung, Frank J Lovicu, and Juergen K V Reichardt. Current progress in using vitamin d and its analogs for cancer prevention and treatment. *Expert Rev Anticancer Ther*, 12(6):811–37, Jun 2012.

[31] V. Cleghon and D. K. Morrison. Raf-1 interacts with fyn and src in a non-phosphotyrosine-dependent manner. *J Biol Chem*, 269(26):17749–17755, Jul 1994.

[32] J. Congleton, H. Jiang, F. Malavasi, H. Lin, and A. Yen. ATRA-induced HL-60 myeloid leukemia cell differentiation depends on the CD38 cytosolic tail needed for membrane localization, but CD38 enzymatic activity is unnecessary. *Exp Cell Res*, 317:910–919, 2011.

[33] J. Congleton, R. MacDonald, and A. Yen. Src inhibitors, PP2 and dasatinib, increase retinoic acid-induced association of Lyn and c-Raf (S259) and enhance MAPK-dependent differentiation of myeloid leukemia cells. *Leukemia*, 26:1180–8, 2012.

[34] Johanna Congleton, Robert MacDonald, and Andrew Yen. Src inhibitors, pp2 and dasatinib, increase retinoic acid-induced association of lyn and c-raf (s259) and enhance mapk-dependent differentiation of myeloid leukemia cells. *Leukemia*, 26(6):1180–1188, 2012.

[35] Johanna Congleton, Miaoqing Shen, Robert MacDonald, Fabio Malavasi, and Andrew Yen. Phosphorylation of c-cbl and p85 pi3k driven by all-trans retinoic acid and cd38 depends on lyn kinase activity. *Cellular signalling*, 26(7):1589–1597, 2014.

[36] Robert J Conrado, Thomas J Mansell, Jeffrey D Varner, and Matthew P DeLisa. Stochastic reaction–diffusion simulation of enzyme compartmentalization reveals improved catalytic efficiency for a synthetic metabolic pathway. *Metabolic engineering*, 9(4):355–363, 2007.

[37] C C Coombs, M Tavakkoli, and M S Tallman. Acute promyelocytic leukemia: where did we start, where are we now, and the future. *Blood Cancer J*, 5:e304, Apr 2015.

[38] CC Coombs, M Tavakkoli, and MS Tallman. Acute promyelocytic leukemia: where did we start, where are we now, and the future. *Blood cancer journal*, 5(4):e304, 2015.

[39] Jorge E Cortes, Hagop M Kantarjian, Tim H Brümmendorf, Dong-Wook Kim, Anna G Turkina, Zhi-Xiang Shen, Ricardo Pasquini, H Jean Khoury, Steven Arkin, Angela Volkert, et al. Safety and efficacy of bosutinib (ski-606) in chronic phase philadelphia chromosome–positive chronic myeloid leukemia patients with resistance or intolerance to imatinib. *Blood*, 118(17):4567–4576, 2011.

[40] P. S. Costello, A. E. Walters, P. J. Mee, M. Turner, L. F. Reynolds, A. Prisco, N. Sarner, R. Zamoyska, and V. L. Tybulewicz. The rho-family gtp exchange factor vav is a critical transducer of t cell receptor signals to the calcium, erk, and nf-kappab pathways. *Proc Natl Acad Sci U S A*, 96(6):3035–3040, Mar 1999.

[41] Markus W Covert and Bernhard Ø Palsson. Transcriptional regulation in constraints-based metabolic models of escherichia coli. *Journal of Biological Chemistry*, 277(31):28058–28064, 2002.

[42] Richard Dahl, Sangeeta R Iyer, Kristin S Owens, Dorothy D Cuylear, and M Celeste Simon. The transcriptional repressor gfi-1 antagonizes pu.1 activity through protein-protein interaction. *J Biol Chem*, 282(9):6473–83, Mar 2007.

[43] Richard Dahl, Jonathan C Walsh, David Lancki, Peter Laslo, Sangeeta R Iyer, Harinder Singh, and M Celeste Simon. Regulation of macrophage and neutrophil cell fates by the PU.1:C/EBPalpha ratio and granulocyte colony-stimulating factor. *Nat Immunol*, 4(10):1029–36, Oct 2003.

[44] Francesco D'Alo', Lisa M Johansen, Erik A Nelson, Hanna S Radomska, Erica K Evans, Pu Zhang, Claus Nerlov, and Daniel G Tenen. The amino terminal and e2f interaction domains are critical for c/ebp alpha-mediated induction of granulopoietic development of hematopoietic cells. *Blood*, 102(9):3163–71, Nov 2003.

[45] James E Darnell, Jr. Reflections on the history of pre-mrna processing and highlights of current knowledge: a unified picture. *RNA*, 19(4):443–60, Apr 2013.

[46] Xavier Darzacq, Yaron Shav-Tal, Valeria de Turris, Yehuda Brody,

Shailesh M Shenoy, Robert D Phair, and Robert H Singer. In vivo dynamics of rna polymerase ii transcription. *Nat Struct Mol Biol*, 14(9):796–806, Sep 2007.

[47] P Delerive, K De Bosscher, S Besnard, W Vanden Berghe, J M Peters, F J Gonzalez, J C Fruchart, A Tedgui, G Haegeman, and B Staels. Peroxisome proliferator-activated receptor alpha negatively regulates the vascular inflammatory gene response by negative cross-talk with transcription factors nf-kappab and ap-1. *J Biol Chem*, 274(45):32048–54, Nov 1999.

[48] Amardeep S Dhillon, Yan Yan Yip, G. Joan Grindlay, Julian L Pakay, Marc Dangers, Meike Hillmann, William Clark, Andrew Pitt, Harald Mischak, and Walter Kolch. The c-terminus of raf-1 acts as a 14-3-3-dependent activation switch. *Cell Signal*, 21(11):1645–1651, Nov 2009.

[49] Joanna R Dispirito, Bin Fang, Fenfen Wang, and Mitchell A Lazar. Pruning of the adipocyte peroxisome proliferator-activated receptor γ cistrome by hematopoietic master regulator pu.1. *Mol Cell Biol*, 33(16):3354–64, Aug 2013.

[50] Mary K Doherty, Dean E Hammond, Michael J Clague, Simon J Gaskell, and Robert J Beynon. Turnover of the human proteome: determination of protein intracellular stability by dynamic silac. *J Proteome Res*, 8(1):104–12, Jan 2009.

[51] MM Domach, SK Leung, RE Cahn, GG Cocks, and ML Shuler. Computer model for glucose-limited growth of a single cell of escherichia coli b/r-a. *Biotechnology and bioengineering*, 26(3):203–216, 1984.

[52] Cédric Dos Santos, Cécile Demur, Valérie Bardet, Nais Prade-Houdellier, Bernard Payrastre, and Christian Récher. A critical role for lyn in acute myeloid leukemia. *Blood*, 111(4):2269–2279, 2008.

[53] J Drach, T McQueen, H Engel, M Andreeff, K A Robertson, S J Collins, F Malavasi, and K Mehta. Retinoic acid-induced expression of cd38 antigen in myeloid cells is mediated through retinoic acid receptor-alpha. *Cancer Res*, 54(7):1746–52, Apr 1994.

[54] Zhijun Duan and Marshall Horwitz. Targets of the transcriptional repressor oncoprotein gfi-1. *Proc Natl Acad Sci U S A*, 100(10):5932–7, May 2003.

[55] Natalie C Duarte, Scott A Becker, Neema Jamshidi, Ines Thiele, Monica L Mo, Thuy D Vo, Rohith Srivas, and Bernhard Ø Palsson. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proceedings of the National Academy of Sciences*, 104(6):1777–1782, 2007.

[56] N.D. Duarte, S.A. Becker, N. Jamshidi, I. Thiele, M.L. Mo, T.D. Vo, R. Srivas, and B.O. Palsson. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci U S A*, 104:1777 – 1782, 2007.

[57] Nicolas Dumaz and Richard Marais. Protein kinase a blocks raf-1 activity by stimulating 14-3-3 binding and blocking raf-1 interaction with ras. *Journal of Biological Chemistry*, 278(32):29819–29823, 2003.

[58] J S Edwards and B Ø Palsson. The escherichia coli mg1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proc Natl Acad Sci U S A*, 97(10):5528–33, May 2000.

[59] Jeremy S Edwards, Rafael U Ibarra, and Bernhard O Palsson. In silico predictions of escherichia coli metabolic capabilities are consistent with experimental data. *Nature biotechnology*, 19(2):125–130, 2001.

[60] Jeremy S Edwards and Bernhard O Palsson. Robustness analysis of the escherichiacoli metabolic network. *Biotechnology Progress*, 16(6):927–939, 2000.

[61] Jeremy S Edwards, Ramprasad Ramakrishna, and Bernhard O Palsson. Characterizing the metabolic phenotype: a phenotype phase plane analysis. *Biotechnology and bioengineering*, 77(1):27–36, 2002.

[62] JS Edwards and BO Palsson. The escherichia coli mg1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences*, 97(10):5528–5533, 2000.

[63] Jame El-Benna, Pham My-Chan Dang, Marie Anne Gougerot-Pocidalo, Jean Claude Marie, and Francoise Braut-Boucher. p47phox, the phagocyte nadph oxidase/nox2 organizer: structure, phosphorylation and implication in diseases. *Exp Mol Med*, 41(4):217–25, Apr 2009.

[64] Philipp Erdrich, Ralf Steuer, and Steffen Klamt. An algorithm for the reduction of genome-scale metabolic network models to meaningful core models. *BMC systems biology*, 9(1):48, 2015.

[65] Valeria R Fantin, Julie St-Pierre, and Philip Leder. Attenuation of ldh-a expression uncovers a link between glycolysis, mitochondrial physiology, and tumor maintenance. *Cancer cell*, 9(6):425–434, 2006.

[66] Jia Fei, Carla Cook, Miriah Gillespie, Bangning Yu, Khyra Fullen, and Nalini Santanam. Atherogenic ω-6 lipids modulate ppar- egr-1 crosstalk in vascular cells. *PPAR Res*, 2011:753917, 2011.

[67] Adam M Feist, Christopher S Henry, Jennifer L Reed, Markus Krummenacker, Andrew R Joyce, Peter D Karp, Linda J Broadbelt, Vassily Hatzimanikatis, and Bernhard Ø Palsson. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol*, 3:121, 2007.

[68] Adam M Feist, Christopher S Henry, Jennifer L Reed, Markus Krummenacker, Andrew R Joyce, Peter D Karp, Linda J Broadbelt, Vassily Hatzimanikatis, and Bernhard Ø Palsson. A genome-scale metabolic reconstruction for escherichia coli k-12 mg1655 that accounts for 1260 orfs and thermodynamic information. *Molecular systems biology*, 3(1), 2007.

[69] Adam M Feist, Markus J Herrgrd, Ines Thiele, Jennie L Reed, and Bernhard Ø Palsson. Reconstruction of biochemical networks in microorganisms. *Nature Reviews Microbiology*, 7(2):129–143, 2009.

[70] Adam M Feist, Markus J Herrgrd, Ines Thiele, Jennie L Reed, and Bernhard Ø Palsson. Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol*, 7(2):129–43, Feb 2009.

[71] David Fell and Athel Cornish-Bowden. *Understanding the control of metabolism*, volume 2. Portland press London, 1997.

[72] J.E. Ferrell. Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr Opin Cell Biol*, 14:140–8, 2002.

[73] Sarah J Freemantle, Michael J Spinella, and Ethan Dmitrovsky. Retinoids in cancer therapy and chemoprevention: promise meets resistance. *Oncogene*, 22(47):7305–15, Oct 2003.

[74] Sarah J Freemantle, Michael J Spinella, and Ethan Dmitrovsky. Retinoids in cancer therapy and chemoprevention: promise meets resistance. *Oncogene*, 22(47):7305, 2003.

[75] R Freitas and R Merkle. *Kinematic Self-Replicating Machines*. Oxford University Press, 2004.

[76] A D Friedman. Transcriptional control of granulocyte and monocyte development. *Oncogene*, 26(47):6816–28, Oct 2007.

[77] Mingui Fu, Jifeng Zhang, Yiming Lin, Xiaojun Zhu, Markus U Ehrengruber, and Yuqing E Chen. Early growth response factor-1 is a critical transcriptional mediator of peroxisome proliferator-activated receptor-gamma 1 gene expression in human aortic smooth muscle cells. *J Biol Chem*, 277(30):26808–14, Jul 2002.

[78] Mingui Fu, Jifeng Zhang, Yiming Lin, Xiaojun Zhu, Markus U Ehrengruber, and Yuqing E Chen. Early growth response factor-1 is a critical transcriptional mediator of peroxisome proliferator-activated receptor-γ1 gene expression in human aortic smooth muscle cells. *Journal of Biological Chemistry*, 277(30):26808–26814, 2002.

[79] Jonathan Garamella, Ryan Marshall, Mark Rustad, and Vincent Noireaux. The all e. coli tx-tl toolbox 2.0: A platform for cell-free synthetic biology. *ACS Synth Biol*, 5(4):344–55, Apr 2016.

[80] Wendy M Geil and Andrew Yen. Nuclear raf-1 kinase regulates the cxcr5 promoter by associating with nfatc3 to drive retinoic acid-induced leukemic cell differentiation. *FEBS J*, 281(4):1170–80, Feb 2014.

[81] Jose A Gomez, Kai Höffner, and Paul I Barton. Dfbalab: a fast and reliable matlab code for dynamic flux balance analysis. *BMC Bioinformatics*, 15:409, Dec 2014.

[82] Jose Alberto Gomez and Paul I Barton. Dynamic flux balance analysis using dfbalab. *Methods Mol Biol*, 1716:353–370, 2018.

[83] Russell Gould, David M Bassen, Anirikh Chakrabarti, Jeffrey D Varner, and Jonathan Butcher. Population heterogeneity in the epithelial to mesenchymal transition is controlled by nfat and phosphorylated sp1. *PLoS Comput Biol*, 12(12):e1005251, Dec 2016.

[84] D. Graham, C. Robertson, J. Bautista, F. Mascarenhas, M.J. Diacovo, V. Montgrain, S-K. Lam, V. Cremasco, W.M. Dunne, R. Faccio, C. Coopersmith, and W. Swat. Neutrophil-mediated oxidative burst and host defense are controlled by a Vav-PLCgamma2 signaling axis in mice. *J Clin Invest*, 117:3445–3452, 2007.

[85] Joshua J Hamilton, Vivek Dwivedi, and Jennifer L Reed. Quantitative Assessment of Thermodynamic Constraints on the Solution Space of Genome-Scale Metabolic Models. *Biophys J*, 105(2):512–522, Jul 2013.

[86] Shouwei Han, Neil Sidell, Paul B Fisher, and Jesse Roman. Up-regulation of p21 gene expression by peroxisome proliferator-activated receptor gamma in human lung carcinoma cells. *Clin Cancer Res*, 10(6):1911–9, Mar 2004.

[87] Douglas Hanahan and Robert A Weinberg. The hallmarks of cancer. *cell*, 100(1):57–70, 2000.

[88] Timothy J Hanly and Michael A Henson. Dynamic flux balance modeling of microbial co-cultures for efficient batch fermentation of glucose and xylose mixtures. *Biotechnol Bioeng*, 108(2):376–85, Feb 2011.

[89] T E Harris, J H Albrecht, M Nakanishi, and G J Darlington. Ccaat/enhancer-binding protein-alpha cooperates with p21 to inhibit cyclin-dependent kinase-2 activity and induces growth arrest independent of dna binding. *J Biol Chem*, 276(31):29200–9, Aug 2001.

[90] Jonathan S Harrison, Xuening Wang, and George P Studzinski. The role of vdr and bim in potentiation of cytarabine-induced cell death in human aml blasts. *Oncotarget*, 7(24):36447–36460, Jun 2016.

[91] Mirko Hekman, Stefan Wiese, Renate Metz, Stefan Albert, Jakob Troppmair, Joachim Nickel, Michael Sendtner, and Ulf R Rapp. Dynamic changes in c-raf phosphorylation and 14-3-3 protein binding in response to growth factor stimulation: differential roles of 14-3-3 protein binding sites. *J Biol Chem*, 279(14):14074–14086, Apr 2004.

[92] Christopher S Henry, Linda J Broadbelt, and Vassily Hatzimanikatis. Thermodynamics-Based Metabolic Flux Analysis. *Biophys. J*, 92(5):192–1805, Mar 2006.

[93] Michael A Henson and Timothy J Hanly. Dynamic flux balance analysis

for synthetic microbial communities. *IET Syst Biol*, 8(5):214–29, Oct 2014.

[94] D. D. Hickstein, A. L. Back, and S. J. Collins. Regulation of expression of the cd11b and cd18 subunits of the neutrophil adherence receptor during human myeloid differentiation. *J Biol Chem*, 264(36):21812–21817, Dec 1989.

[95] Hirohumi Hirayama, Kiyono Yoshii, Hidetomo Ojima, Norikazu Kawai, Shintaro Gotoh, and Yuzo Fukuyama. Linear systems analysis of activating processes of complement system as a defense mechanism. *Biosystems*, 39(3):173–185, 1996.

[96] J L Hjersted and M A Henson. Steady-state and dynamic flux balance analysis of ethanol production by saccharomyces cerevisiae. *IET Syst Biol*, 3(3):167–79, May 2009.

[97] Jared L Hjersted, Michael A Henson, and Radhakrishnan Mahadevan. Genome-scale analysis of saccharomyces cerevisiae metabolism and ethanol production in fed-batch culture. *Biotechnol Bioeng*, 97(5):1190–204, Aug 2007.

[98] Hanno Hock, Melanie J Hamblen, Heather M Rooke, David Traver, Roderick T Bronson, Scott Cameron, and Stuart H Orkin. Intrinsic requirement for zinc finger transcription factor gfi-1 in neutrophil differentiation. *Immunity*, 18(1):109–20, Jan 2003.

[99] C Eric Hodgman and Michael C Jewett. Cell-free synthetic biology: thinking outside the cell. *Metab Eng*, 14(3):261–9, May 2012.

[100] K Höffner, S M Harwood, and P I Barton. A reliable simulator for dynamic flux balance analysis. *Biotechnol Bioeng*, 110(3):792–802, Mar 2013.

[101] H. Y. Hong, S. Varvayanis, and A. Yen. Retinoic acid causes MEK-dependent RAF phosphorylation through RARalpha plus RXR activation in HL-60 cells. *Differentiation*, 68:55–66, 2001.

[102] I. Hornstein, A. Alcover, and S. Katzav. Vav proteins, masters of the world of cytoskeleton organization. *Cell Signal*, 16:1–11, 2004.

[103] Nicholas Horvath, Michael Vilkhovoy, Joseph A. Wayman, Kara Cal-

houn, James Swartz, and Jeffrey Varner. Toward a genome scale sequence specific dynamic model of cell-free protein synthesis in escherichia coli. *bioRxiv*, 2017.

[104] Chelsea Y Hu, Jeffrey D Varner, and Julius B Lucks. Generating effective models and parameters for rna genetic circuits. *ACS Synth Biol*, 4(8):914–26, Aug 2015.

[105] Sophia Hu, Masumi Ueda, Lindsay Stetson, James Ignatz-Hoover, Stephen Moreton, Amit Chakrabarti, Zhiqiang Xia, Goutam Karan, Marcos de Lima, Mukesh K Agrawal, et al. A novel glycogen synthase kinase-3 inhibitor optimized for acute myeloid leukemia differentiation activity. *Molecular cancer therapeutics*, 15(7):1485–1494, 2016.

[106] Daniel R Hyduke, Nathan E Lewis, and Bernhard Ø Palsson. Analysis of omics data with genome-scale models of metabolism. *Mol Biosyst*, 9(2):167–74, Feb 2013.

[107] Christopher N Ibabao, Rodica P Bunaciu, Deanna MW Schaefer, and Andrew Yen. The ahr agonist vaf347 augments retinoic acid-induced differentiation in leukemia cells. *FEBS open bio*, 5(1):308–318, 2015.

[108] Rafael U Ibarra, Jeremy S Edwards, and Bernhard O Palsson. Escherichia coli k-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature*, 420(6912):186–189, 2002.

[109] Rafael U Ibarra, Jeremy S Edwards, and Bernhard Ø Palsson. Escherichia coli k-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature*, 420(6912):186–9, Nov 2002.

[110] Christian Igel, Nikolaus Hansen, and Stefan Roth. Covariance matrix adaptation for multi-objective optimization. *Evol Comput*, 15(1):1–28, 2007.

[111] Noriyoshi Iriyama, Bo Yuan, Yoshihiro Hatta, Norio Takagi, and Masami Takei. Lyn, a tyrosine kinase closely linked to the differentiation status of primary acute myeloid leukemia blasts, associates with negative regulation of all-trans retinoic acid (atra) and dihydroxyvitamin d3 (vd3)-induced hl-60 cells differentiation. *Cancer Cell International*, 16(1):37, 2016.

[112] D A Jackson, A Pombo, and F Iborra. The balance sheet for tran-

scription: an analysis of nuclear rna metabolism in mammalian cells. *FASEB J*, 14(2):242–54, Feb 2000.

[113] Thapakorn Jaroentomeechai, Jessica C Stark, Aravind Natarajan, Cameron J Glasscock, Laura E Yates, Karen J Hsu, Milan Mrksich, Michael C Jewett, and Matthew P DeLisa. Single-pot glycoprotein biosynthesis using a cell-free transcription-translation system enriched with glycosylation machinery. *Nat Commun*, 9(1):2686, Jul 2018.

[114] Holly A Jensen, Rodica P Bunaciu, Christopher N Ibabao, Rebecca Myers, Jeffrey D Varner, and Andrew Yen. Retinoic acid therapy resistance progresses from unilineage to bilineage in hl-60 leukemic blasts. *PLoS One*, 9(6):e98929, 2014.

[115] Holly A Jensen, Rodica P Bunaciu, Christopher N Ibabao, Rebecca Myers, Jeffrey D Varner, and Andrew Yen. Retinoic acid therapy resistance progresses from unilineage to bilineage in hl-60 leukemic blasts. *PloS one*, 9(6):e98929, 2014.

[116] Holly A Jensen, Rodica P Bunaciu, Jeffrey D Varner, and Andrew Yen. Gw5074 and pp2 kinase inhibitors implicate nontraditional c-raf and lyn function as drivers of retinoic acid-induced maturation. *Cell Signal*, 27(8):1666–75, Aug 2015.

[117] Holly A Jensen, Lauren E Styskal, Ryan Tasseff, Rodica P Bunaciu, Johanna Congleton, Jeffrey D Varner, and Andrew Yen. The src-family kinase inhibitor pp2 rescues inducible differentiation events in emergent retinoic acid-resistant myeloblastic leukemia cells. *PLoS One*, 8(3):e58621, 2013.

[118] Holly A Jensen, Lauren E Styskal, Ryan Tasseff, Rodica P Bunaciu, Johanna Congleton, Jeffrey D Varner, and Andrew Yen. The src-family kinase inhibitor pp2 rescues inducible differentiation events in emergent retinoic acid-resistant myeloblastic leukemia cells. *PLoS One*, 8(3):e58621, 2013.

[119] Holly A Jensen, Harmony B Yourish, Rodica P Bunaciu, Jeffrey D Varner, and Andrew Yen. Induced myelomonocytic differentiation in leukemia cells is accompanied by noncanonical transcription factor expression. *FEBS Open Bio*, 5:789–800, 2015.

[120] M.C. Jewett, A. Voloshin, and J. Swartz. *Prokaryotic systems for in*

vitro expression, pages 391–411. Eaton Publishing, Westborough, MA, 2002.

[121] Michael C Jewett, Kara A Calhoun, Alexei Voloshin, Jessica J Wuu, and James R Swartz. An integrated cell-free metabolic platform for protein production and synthetic biology. *Mol Syst Biol*, 4:220, 2008.

[122] Michael C. Jewett and James R. Swartz. Mimicking the escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. *Biotechnology and Bioengineering*, 86(1):19–26, 2004.

[123] D G Johnson, K Ohtani, and J R Nevins. Autoregulatory control of e2f1 expression in response to positive and negative regulators of cell cycle progression. *Genes Dev*, 8(13):1514–25, Jul 1994.

[124] Hagop M Kantarjian, Jorge E Cortes, Dong-Wook Kim, H Jean Khoury, Tim H Brümmendorf, Kimmo Porkka, Giovanni Martinelli, Simon Durrant, Eric Leip, Virginia Kelly, et al. Bosutinib safety and management of toxicity in leukemia patients with resistance or intolerance to imatinib and other tyrosine kinase inhibitors. *Blood*, 123(9):1309–1318, 2014.

[125] D Kardassis, P Papakosta, K Pardali, and A Moustakas. c-jun transactivates the promoter of the human p21(waf1/cip1) gene by acting as a superactivator of the ubiquitous transcription factor sp1. *J Biol Chem*, 274(41):29572–81, Oct 1999.

[126] K. Katagiri, S. Hattori, S. Nakamura, T. Yamamoto, T. Yoshida, and T. Katagiri. Activation of ras and formation of gap complex during tpa-induced monocytic differentiation of hl-60 cells. *Blood*, 84(6):1780–1789, Sep 1994.

[127] Koko Katagiri, Seisuke Hattori, S Nakamura, T Yamamoto, Takeshi Yoshida, and Takuya Katagiri. Activation of ras and formation of gap complex during tpa-induced monocytic differentiation of hl-60 cells. *Blood*, 84(6):1780–1789, 1994.

[128] Koko Katagiri, Kazunari K Yokoyama, Tadashi Yamamoto, Satoshi Omura, Shinkichi Irie, and Takuya Katagiri. Lyn and fgr protein-tyrosine kinases prevent apoptosis during retinoic acid-induced granulocytic differentiation of hl-60 cells. *Journal of Biological Chemistry*, 271(19):11557–11562, 1996.

[129] Ryan Tasseff Caitlin Gee Matthew P. DeLisa Jeffrey D. Varner Katharine V. Rogers, Joseph A. Wayman. Modeling and analysis of hormone and mitogenic signal integration in prostate cancer. *BioArxive*, 2016.

[130] Kenneth J Kauffman, Purusharth Prakash, and Jeremy S Edwards. Advances in flux balance analysis. *Current opinion in biotechnology*, 14(5):491–496, 2003.

[131] Boris Kholodenko, Michael B Yaffe, and Walter Kolch. Computational approaches for analyzing information flow in biological networks. *Sci. Signal.*, 5(220):re1–re1, 2012.

[132] Boris N Kholodenko, Oleg V Demin, Gisela Moehren, and Jan B Hoek. Quantification of short term signaling by the epidermal growth factor receptor. *Journal of Biological Chemistry*, 274(42):30169–30181, 1999.

[133] Takanori Kigawa, Yutaka Muto, and Shigeyuki Yokoyama. Cell-free synthesis and amino acid-selective stable isotope labeling of proteins for NMR analysis. *J Biomolec NMR*, 6(2):129–134, 1995.

[134] Dong-Myung Kim and James R. Swartz. Regeneration of adenosine triphosphate from glycolytic intermediates for cell-free protein synthesis. *Biotech. Bioeng.*, 74(4):309–316, 2001.

[135] Hong Seok Kim and In Kyoung Lim. Phosphorylated extracellular signal-regulated protein kinases 1 and 2 phosphorylate sp1 on serine 59 and regulate cellular senescence via transcription of p21sdi1/cip1/waf1. *J Biol Chem*, 284(23):15475–15486, Jun 2009.

[136] MP Kim, SI Park, S Kopetz, and Gary E Gallick. Src family kinases as mediators of endothelial permeability: effects on inflammation and metastasis. *Cell and tissue research*, 335(1):249, 2009.

[137] Walter Kolch, Gisela Heidecker, Georg Kochs, Richard Hummel, Haleh Vahidi, Harald Mischak, Günter Finkenzeller, Dieter Marmé, and Ulf R Rapp. Protein kinase cα activates raf-1 by direct phosphorylation. *Nature*, 364(6434):249–252, 1993.

[138] Dhinakar S Kompala, Doraiswami Ramkrishna, Norman B Jansen, and George T Tsao. Investigation of bacterial growth on mixed substrates: experimental evaluation of cybernetic models. *Biotechnology and Bioengineering*, 28(7):1044–1055, 1986.

[139] Andrey A Korotaevskiy, Leonid G Hanin, and Mikhail A Khanin. Nonlinear dynamics of the complement system activation. *Math Biosci.*, 222(2):127–143, 2009.

[140] Martin Kos and David Tollervey. Yeast pre-rrna processing and modification occur cotranscriptionally. *Mol Cell*, 37(6):809–20, Mar 2010.

[141] PL Kropf, L Wang, Y Zang, RL Redner, and DE Johnson. Dasatinib promotes atra-induced differentiation of aml cells. *Leukemia*, 24(3):663, 2010.

[142] Peter Laslo, Chauncey J Spooner, Aryeh Warmflash, David W Lancki, Hyun-Jun Lee, Roger Sciammas, Benjamin N Gantner, Aaron R Dinner, and Harinder Singh. Multilineage transcriptional priming and determination of alternate hematopoietic cell fates. *Cell*, 126(4):755–66, Aug 2006.

[143] Sun Bok Lee and James E. Bailey. Genetically structured models forlac promoter–operator function in the escherichia coli chromosome and in multicopy plasmids: Lac operator function. *Biotechnology and Bioengineering*, 26(11):1372–1382, 1984.

[144] Joshua Lequieu, Anirikh Chakrabarti, Satyaprakash Nayak, and Jeffrey D Varner. Computational modeling and analysis of insulin induced eukaryotic translation initiation. *PLoS computational biology*, 7(11):e1002263, 2011.

[145] Brian C Lewis, Hyunsuk Shim, Qing Li, Chyi Sun Wu, Linda A Lee, Amit Maity, and Chi V Dang. Identification of putative c-myc-responsive genes: characterization of rcl, a novel growth-related gene. *Molecular and cellular biology*, 17(9):4967–4978, 1997.

[146] Nathan E Lewis, Harish Nagarajan, and Bernhard O Palsson. Constraining the metabolic genotype–phenotype relationship using a phylogeny of in silico methods. *Nature Reviews Microbiology*, 10(4):291–305, 2012.

[147] Nathan E Lewis, Harish Nagarajan, and Bernhard Ø Palsson. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat Rev Microbiol*, 10(4):291–305, Apr 2012.

[148] Chuan Li, Masahiko Imai, Shinya Hasegawa, Masahiro Yamasaki, and Noriko Takahashi. Growth inhibition of refractory human gallbladder cancer cells by retinol, and its mechanism of action. *Biol Pharm Bull*, 40(4):495–503, Apr 2017.

[149] S L Li, W Schlegel, A J Valente, and R A Clark. Critical flanking sequences of pu.1 binding sites in myeloid-specific promoters. *J Biol Chem*, 274(45):32453–60, Nov 1999.

[150] Xinran Li, Cen Guo, Yanan Li, Lina Li, Yuxin Wang, Yiming Zhang, Yue Li, Yu Chen, Wenhan Liu, and Li Gao. Ketamine administered pregnant rats impair learning and memory in offspring via the creb pathway. *Oncotarget*, Feb 2017.

[151] Zegui Li, Michelle Hedrick Theus, and Ling Wei. Role of erk 1/2 signaling in neuronal differentiation of cultured embryonic stem cells. *Dev Growth Differ*, 48(8):513–523, Oct 2006.

[152] James C Liao. Modelling and analysis of metabolic pathways. *Current opinion in biotechnology*, 4(2):211–216, 1993.

[153] Maria Rosa Lidonnici, Alessandra Audia, Angela Rachele Soliera, Marco Prisco, Giovanna Ferrari-Amorotti, Todd Waldron, Nick Donato, Ying Zhang, Robert V Martinez, Tessa L Holyoake, and Bruno Calabretta. Expression of the transcriptional repressor gfi-1 is regulated by c/ebpalpha and is involved in its proliferation and colony formation-inhibitory effects in p210bcr/abl-expressing cells. *Cancer Res*, 70(20):7949–59, Oct 2010.

[154] Bing Liu, Jing Zhang, Pei Yi Tan, David Hsu, Anna M Blom, Benjamin Leong, Sunil Sethi, Bow Ho, Jeak Ling Ding, and PS Thiagarajan. A computational and experimental study of the regulatory mechanisms of the complement system. *PLoS Comput Biol*, 7(1):e1001059, 2011.

[155] M Liu, A Iavarone, and L P Freedman. Transcriptional activation of the human p21(waf1/cip1) gene by retinoic acid receptor. correlation with retinoid induction of u937 cell differentiation. *J Biol Chem*, 271(49):31723–8, Dec 1996.

[156] Jeremy S Logue and Deborah K Morrison. Complexity in the signaling network: insights from the use of targeted inhibitors in cancer therapy. *Genes & development*, 26(7):641–650, 2012.

[157] Yuan Lu, John P Welsh, and James R Swartz. Production and stabilization of the trimeric influenza hemagglutinin stem domain for potentially broadly protective influenza vaccines. *Proc Natl Acad Sci U S A*, 111(1):125–30, Jan 2014.

[158] Deyan Luan, Michael Zai, and Jeffrey D Varner. Computationally derived points of fragility of a human cascade are consistent with current therapeutic strategies. *PLoS Comput Biol*, 3(7):e142, Jul 2007.

[159] Weibo Luo, Hongxia Hu, Ryan Chang, Jun Zhong, Matthew Knabel, Robert O'Meally, Robert N Cole, Akhilesh Pandey, and Gregg L Semenza. Pyruvate kinase m2 is a phd3-stimulated coactivator for hypoxia-inducible factor 1. *Cell*, 145(5):732–744, 2011.

[160] Xin M Luo and A Catharine Ross. Retinoic acid exerts dual regulatory actions on the expression and nuclear localization of interferon regulatory factor-1. *Exp Biol Med (Maywood)*, 231(5):619–31, May 2006.

[161] Radhakrishnan Mahadevan, Jeremy S Edwards, and Francis J Doyle. Dynamic flux balance analysis of diauxic growth in escherichia coli. *Biophysical journal*, 83(3):1331–1340, 2002.

[162] Radhakrishnan Mahadevan, Jeremy S Edwards, and Francis J Doyle, 3rd. Dynamic flux balance analysis of diauxic growth in escherichia coli. *Biophys J*, 83(3):1331–40, Sep 2002.

[163] Ka Sin Mak, Alister P W Funnell, Richard C M Pearson, and Merlin Crossley. Pu.1 and haematopoietic cell fate: Dosage matters. *Int J Cell Biol*, 2011:808524, 2011.

[164] D. J. Mangelsdorf, E. S. Ong, J. A. Dyck, and R. M. Evans. Nuclear receptor that identifies a novel retinoic acid response pathway. *Nature*, 345:224–229, 1990.

[165] M Marchisio, V Bertagnolo, M L Colamussi, S Capitani, and L M Neri. Phosphatidylinositol 3-kinase in HL-60 nuclei is bound to the nuclear matrix and increases during granulocytic differentiation. *Biochem Biophys Res Commun*, 253:346–51, 1998.

[166] Tamiko Matsuda, Tamio Noguchi, Masaru Takenaka, Kazuya Yamada, and Takehiko Tanaka. Regulation of l-type pyruvate kinase gene expression by dietary fructose in normal and diabetic rats. *The Journal of Biochemistry*, 107(4):655–660, 1990.

[167] J H Matthaei and M W Nirenberg. Characteristics and stabilization of dnaase-sensitive protein synthesis in e. coli extracts. *Proc Natl Acad Sci U S A*, 47:1580–8, Oct 1961.

[168] Douglas McCloskey, Bernhard Ø Palsson, and Adam M Feist. Basic and applied uses of genome-scale metabolic network reconstructions of escherichia coli. *Mol Syst Biol*, 9:661, Apr 2013.

[169] R Meyers, editor. *Encyclopedia of Molecular Cell Biology and Molecular Medicine, Volume 1, 2nd Edition*. Number ISBN: 978-3-527-30543-8. Wiley-Blackwell, 2004.

[170] Ping Miao, Shile Sheng, Xiaoguang Sun, Jianjun Liu, and Gang Huang. Lactate dehydrogenase a in cancer: a promising target for diagnosis and therapy. *IUBMB life*, 65(11):904–910, 2013.

[171] Julie Milanini-Mongiat, Jacques Pouysségur, and Gilles Pagès. Identification of two sp1 phosphorylation sites for p42/p44 mitogen-activated protein kinases: their implication in vascular endothelial growth factor gene transcription. *J Biol Chem*, 277(23):20631–20639, Jun 2002.

[172] Ron Milo, Paul Jorgensen, Uri Moran, Griffin Weber, and Michael Springer. Bionumbers–the database of key numbers in molecular and cell biology. *Nucleic Acids Res*, 38(Database issue):D750–3, Jan 2010.

[173] M. B. Miranda and D. E. Johnson. Signal transduction pathways that contribute to myeloid differentiation. *Leukemia*, 21(7):1363–1377, Jul 2007.

[174] Tae Seok Moon, Chunbo Lou, Alvin Tamsir, Brynne C Stanton, and Christopher A Voigt. Genetic programs constructed from layered logic gates in single cells. *Nature*, 491(7423):249–53, Nov 2012.

[175] Beatrice U Mueller, Thomas Pabst, José Fos, Vibor Petkovic, Martin F Fey, Norio Asou, Ulrich Buergi, and Daniel G Tenen. Atra resolves the differentiation block in t(15;17) acute myeloid leukemia by restoring pu.1 expression. *Blood*, 107(8):3330–8, Apr 2006.

[176] Arne C Müller and Alexander Bockmayr. Fast thermodynamically constrained flux variability analysis. *Bioinformatics*, 29(7):903–9, Apr 2013.

[177] Toshikage Nagao, Tetsuya Kurosu, Yoshihiro Umezawa, Ayako Nogami, Gaku Oshikawa, Shuji Tohda, Masahide Yamamoto, and Osamu Miura. Proliferation and survival signaling from both jak2-v617f and lyn involving gsk3 and mtor/p70s6k/4ebp1 in pvtl-1 cell line newly established from acute myeloid leukemia transformed from polycythemia vera. *PLoS One*, 9(1):e84746, 2014.

[178] B Nilsson. Probable in vivo induction of differentiation by retinoic acid of promyelocytes in acute promyelocytic leukaemia. *Br J Haematol*, 57(3):365–71, Jul 1984.

[179] M W Nirenberg and J H Matthaei. The dependence of cell-free protein synthesis in e. coli upon naturally occurring or synthetic polyribonucleotides. *Proc Natl Acad Sci U S A*, 47:1588–602, Oct 1961.

[180] Béla Novák, Attila Tóth, Attila Csikász-Nagy, Béla Györffy, John J Tyson, and Kim Nasmyth. Finishing the cell cycle. *Journal of Theoretical Biology*, 199(2):223–233, 1999.

[181] Edward J O'Brien, Joshua A Lerman, Roger L Chang, Daniel R Hyduke, and Bernhard Ø Palsson. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol. Sys. Biol.*, 9(1):693, 2013.

[182] You-Kwan Oh, Bernhard O Palsson, Sung M Park, Christophe H Schilling, and Radhakrishnan Mahadevan. Genome-scale reconstruction of metabolic network in bacillus subtilis based on high-throughput phenotyping and gene essentiality data. *Journal of Biological Chemistry*, 282(39):28791–28799, 2007.

[183] You-Kwan Oh, Bernhard Ø Palsson, Sung M Park, Christophe H Schilling, and Radhakrishnan Mahadevan. Genome-scale reconstruction of metabolic network in bacillus subtilis based on high-throughput phenotyping and gene essentiality data. *J Biol Chem*, 282(39):28791–9, Sep 2007.

[184] Nuala A O'Leary, Mathew W Wright, J Rodney Brister, Stacy Ciufo, Diana Haddad, Rich McVeigh, Bhanu Rajput, Barbara Robbertse, Brian Smith-White, Danso Ako-Adjei, Alexander Astashyn, Azat Badretdin, Yiming Bao, Olga Blinkova, Vyacheslav Brover, Vyacheslav Chetvernin, Jinna Choi, Eric Cox, Olga Ermolaeva, Catherine M Farrell, Tamara Goldfarb, Tripti Gupta, Daniel Haft, Eneida Hatcher, Wratko Hlavina, Vinita S Joardar, Vamsi K Kodali, Wenjun Li, Donna

Maglott, Patrick Masterson, Kelly M McGarvey, Michael R Murphy, Kathleen O'Neill, Shashikant Pujar, Sanjida H Rangwala, Daniel Rausch, Lillian D Riddick, Conrad Schoch, Andrei Shkeda, Susan S Storz, Hanzhen Sun, Francoise Thibaud-Nissen, Igor Tolstoy, Raymond E Tully, Anjana R Vatsan, Craig Wallin, David Webb, Wendy Wu, Melissa J Landrum, Avi Kimchi, Tatiana Tatusova, Michael DiCuccio, Paul Kitts, Terence D Murphy, and Kim D Pruitt. Reference sequence (refseq) database at ncbi: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*, 44(D1):D733–45, Jan 2016.

[185] H L Pahl, R J Scheibe, D E Zhang, H M Chen, D L Galson, R A Maki, and D G Tenen. The proto-oncogene pu.1 regulates expression of the myeloid-specific cd11b promoter. *J Biol Chem*, 268(7):5014–20, Mar 1993.

[186] BO Palsson, A Joshi, and SS Ozturk. Reducing complexity in metabolic networks: making metabolic meshes manageable. In *Federation proceedings*, volume 46, pages 2485–2489, 1987.

[187] Z Pan, C J Hetherington, and D E Zhang. Ccaat/enhancer-binding protein activates the cd14 promoter and mediates transforming growth factor beta signaling in monocyte development. *J Biol Chem*, 274(33):23242–8, Aug 1999.

[188] Keith Pardee, Shimyn Slomovic, Peter Q Nguyen, Jeong Wook Lee, Nina Donghia, Devin Burrill, Tom Ferrante, Fern R McSorley, Yoshikazu Furuta, Andyna Vernet, Michael Lewandowski, Christopher N Boddy, Neel S Joshi, and James J Collins. Portable, on-demand biomolecular manufacturing. *Cell*, 167(1):248–59.e12, Sep 2016.

[189] Toby Passioura, Alla Dolnikov, Sylvie Shen, and Geoff Symonds. N-ras-induced growth suppression of myeloid cells is mediated by irf-1. *Cancer Res*, 65(3):797–804, Feb 2005.

[190] Joseph D Raffetto, Ricardo Vasquez, David G Goodwin, and James O Menzoian. Mitogen-activated protein kinase pathway regulates cell proliferation in venous ulcer fibroblasts. *Vascular and endovascular surgery*, 40(1):59–66, 2006.

[191] Ramprasad Ramakrishna, Jeremy S Edwards, Andrew McCulloch, and Bernhard O Palsson. Flux-balance analysis of mitochondrial energy metabolism: consequences of systemic stoichiometric constraints.

American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, 280(3):R695–R704, 2001.

[192] A K Rishi, T M Gerald, Z M Shao, X S Li, R G Baumann, M I Dawson, and J A Fontana. Regulation of the human retinoic acid receptor alpha gene in the estrogen receptor negative human breast carcinoma cell lines skbr-3 and mda-mb-435. *Cancer Res*, 56(22):5246–52, Nov 1996.

[193] Daniel A Ritt, Ming Zhou, Thomas P Conrads, Timothy D Veenstra, Terry D Copeland, and Deborah K Morrison. Ck2 is a component of the ksr1 scaffold complex that contributes to raf kinase activation. *Curr Biol*, 17(2):179–184, Jan 2007.

[194] David J Roberts, Valerie P Tan-Sah, Jeffery M Smith, and Shigeki Miyamoto. Akt phosphorylates hk-ii at thr-473 and increases mitochondrial hk-ii association to protect cardiomyocytes. *Journal of Biological Chemistry*, 288(33):23798–23806, 2013.

[195] Evan D Rosen, Chung-Hsin Hsu, Xinzhong Wang, Shuichi Sakai, Mason W Freeman, Frank J Gonzalez, and Bruce M Spiegelman. C/ebpalpha induces adipogenesis through ppargamma: a unified pathway. *Genes Dev*, 16(1):22–6, Jan 2002.

[196] Michael J Rosenbluth, Wilbur A Lam, and Daniel A Fletcher. Force microscopy of nonadherent cells: a comparison of leukemia cell deformability. *Biophys J*, 90(8):2994–3003, Apr 2006.

[197] Adithya Sagar, Wei Dai, Mason Minot, Rachel LeCover, and Jeffrey D Varner. Reduced order modeling and analysis of the human complement system. *PloS one*, 12(11):e0187373, 2017.

[198] Adithya Sagar and Jeffrey D Varner. Dynamic modeling of the human coagulation cascade using reduced order effective kinetic models. *Processes*, 3(1):178–203, 2015.

[199] Christophe H Schilling, Markus W Covert, Iman Famili, George M Church, Jeremy S Edwards, and Bernhard O Palsson. Genome-scale metabolic model of helicobacter pylori 26695. *Journal of bacteriology*, 184(16):4582–4593, 2002.

[200] Birgit Schoeberl, Claudia Eichler-Jonsson, Ernst Dieter Gilles, and Gertraud Müller. Computational modeling of the dynamics of the map

kinase cascade activated by surface and internalized egf receptors. *Nature biotechnology*, 20(4):370, 2002.

[201] Robert Schuetz, Lars Kuepfer, and Uwe Sauer. Systematic evaluation of objective functions for predicting intracellular fluxes in escherichia coli. *Mol Syst Biol*, 3:119, 2007.

[202] P B Sehgal, E Derman, G R Molloy, I Tamm, and J E Darnell. 5,6-dichloro-1-beta-d-ribofuranosylbenzimidazole inhibits initiation of nuclear heterogeneous rna chains in hela cells. *Science*, 194(4263):431–3, Oct 1976.

[203] Gregg L Semenza, Bing-Hua Jiang, Sandra W Leung, Rosa Passantino, Jean-Paul Concordet, Pascal Maire, and Agata Giallongo. Hypoxia response elements in the aldolase a, enolase 1, and lactate dehydrogenase a gene promoters contain essential binding sites for hypoxia-inducible factor 1. *Journal of Biological Chemistry*, 271(51):32529–32537, 1996.

[204] Mengrou Shan, David Dai, Arunodai Vudem, Jeffrey D Varner, and Abraham D Stroock. Multi-scale computational study of the warburg effect, reverse warburg effect and glutamine addiction in solid tumors. *PLoS computational biology*, 14(12):e1006584, 2018.

[205] M. Shen, R. Bunaciu, J. Congleton, H. Jensen, L. Sayam, J. Varner, and A. Yen. Interferon regulatory factor-1 binds c-Cbl, enhances mitogen activated protein kinase signaling and promotes retinoic acid-induced differentiation of HL-60 human myelo-monoblastic leukemia cells. *Leuk Lymphoma*, 52:2372–9, 2011.

[206] Miaoqing Shen, Rodica P Bunaciu, Johanna Congleton, Holly A Jensen, Lavanya G Sayam, Jeffrey D Varner, and Andrew Yen. Interferon regulatory factor-1 binds c-cbl, enhances mitogen activated protein kinase signaling and promotes retinoic acid-induced differentiation of hl-60 human myelo-monoblastic leukemia cells. *Leuk Lymphoma*, 52(12):2372–9, Dec 2011.

[207] Miaoqing Shen and Andrew Yen. c-cbl interacts with cd38 and promotes retinoic acid–induced differentiation and g0 arrest of human myeloblastic leukemia cells. *Cancer research*, 68(21):8761–8769, 2008.

[208] Miaoqing Shen and Andrew Yen. c-cbl tyrosine kinase-binding domain mutant g306e abolishes the interaction of c-cbl with cd38 and fails to

promote retinoic acid-induced cell differentiation and g0 arrest. *J Biol Chem*, 284(38):25664–77, Sep 2009.

[209] Hyunsuk Shim, Yoon S Chun, Brian C Lewis, and Chi V Dang. A unique glucose-dependent apoptotic pathway induced by c-myc. *Proceedings of the National Academy of Sciences*, 95(4):1511–1516, 1998.

[210] Tomer Shlomi, Tomer Benyamini, Eyal Gottlieb, Roded Sharan, and Eytan Ruppin. Genome-scale metabolic modeling elucidates the role of proliferative adaptation in causing the warburg effect. *PLoS Comput Biol*, 7(3):e1002018, 2011.

[211] Eun-Kyung Song, Young-Rae Lee, Yu-Ri Kim, Ji-Hyun Yeom, Chae-Hwa Yoo, Hyun-Kag Kim, Hye-Min Park, Hyung-Sub Kang, Jong-Suk Kim, Uh-Hyun Kim, and Myung-Kwan Han. Naadp mediates insulin-stimulated glucose uptake and insulin sensitization by pparγ in adipocytes. *Cell Rep*, 2(6):1607–19, Dec 2012.

[212] J. S. Song, J. Gomez, L. F. Stancato, and J. Rivera. Association of a p95 vav-containing signaling complex with the fcepsilonri gamma chain in the rbl-2h3 mast cell line. evidence for a constitutive in vivo association of vav with grb2, raf-1, and erk2 in an active complex. *J Biol Chem*, 271(43):26962–26970, Oct 1996.

[213] AS Spirin, VI Baranov, LA Ryabova, SY Ovodov, and YB Alakhov. A continuous cell-free translation system capable of producing polypeptides in high yield. *Science*, 242(4882):1162–1164, 1988.

[214] Cindy Starbuck and Douglas A Lauffenburger. Mathematical model for the effects of epidermal growth factor receptor trafficking dynamics on fibroblast proliferation responses. *Biotechnology progress*, 8(2):132–143, 1992.

[215] Ulrich Steidl, Frank Rosenbauer, Roel G W Verhaak, Xuesong Gu, Alexander Ebralidze, Hasan H Otu, Steffen Klippel, Christian Steidl, Ingmar Bruns, Daniel B Costa, Katharina Wagner, Manuel Aivado, Guido Kobbe, Peter J M Valk, Emmanuelle Passegué, Towia A Libermann, Ruud Delwel, and Daniel G Tenen. Essential role of jun family transcription factors in pu.1 knockdown-induced leukemic stem cells. *Nat Genet*, 38(11):1269–77, Nov 2006.

[216] Emma R Still and Mariia O Yuneva. Hopefully devoted to q: targeting

glutamine addiction in cancer. *British journal of cancer*, 116(11):1375, 2017.

[217] Jaehong Suh, Young Jin Jeon, Hwan Mook Kim, Jong Soon Kang, Norbert E Kaminski, and Kyu-Hwan Yang. Aryl hydrocarbon receptor-dependent inhibition of ap-1 activity by 2,3,7,8-tetrachlorodibenzo-p-dioxin in activated b cells. *Toxicol Appl Pharmacol*, 181(2):116–23, Jun 2002.

[218] Gopinath Sutendra, Adam Kinnaird, Peter Dromparis, Roxane Paulin, Trevor H Stenson, Alois Haromy, Kyoko Hashimoto, Nancy Zhang, Eric Flaim, and Evangelos D Michelakis. A nuclear pyruvate dehydrogenase complex is important for the generation of acetyl-coa and histone acetylation. *Cell*, 158(1):84–97, 2014.

[219] Neil Swainston, Kieran Smallbone, Hooman Hefzi, Paul D Dobson, Judy Brewer, Michael Hanscho, Daniel C Zielinski, Kok Siong Ang, Natalie J Gardiner, Jahir M Gutierrez, et al. Recon 2.2: from reconstruction to model of human metabolism. *Metabolomics*, 12(7):109, 2016.

[220] I Sylvester and H R Schöler. Regulation of the oct-4 gene by nuclear receptors. *Nucleic Acids Res*, 22(6):901–11, Mar 1994.

[221] Attila Szanto and Laszlo Nagy. Retinoids potentiate peroxisome proliferator-activated receptor gamma action in differentiation, gene expression, and lipid metabolic processes in developing myeloid cells. *Mol Pharmacol*, 67(6):1935–43, Jun 2005.

[222] Xiao-Han Tang and Lorraine J Gudas. Retinoids, retinoic acid receptors, and cancer. *Annu Rev Pathol*, 6:345–64, 2011.

[223] R. Tasseff, S. Nayak, S.O. Song, A. Yen, and J. Varner. Modeling and analysis of retinoic acid induced differentiation of uncommitted precursor cells. *Integr Biol*, 3:578 – 591, 2011.

[224] Praveena S Thiagarajan, Masahiro Hitomi, James S Hale, Alvaro G Alvarado, Balint Otvos, Maksim Sinyuk, Kevin Stoltz, Andrew Wiechert, Erin Mulkearns-Hubert, Awad M Jarrar, et al. Development of a fluorescent reporter system to delineate cancer stem cells in triple-negative breast cancer. *Stem Cells*, 33(7):2114–2125, 2015.

[225] Ines Thiele, Neema Jamshidi, Ronan MT Fleming, and Bernhard Ø Palsson. Genome-scale reconstruction of escherichia coli's transcriptional and translational machinery: a knowledge base, its mathematical formulation, and its functional characterization. *PLoS Comput Biol*, 5(3):e1000312, 2009.

[226] Ines Thiele, Neil Swainston, Ronan MT Fleming, Andreas Hoppe, Swagatika Sahoo, Maike K Aurich, Hulda Haraldsdottir, Monica L Mo, Ottar Rolfsson, Miranda D Stobbe, et al. A community-driven global reconstruction of human metabolism. *Nature biotechnology*, 31(5):419–425, 2013.

[227] N Timchenko, D R Wilson, L R Taylor, S Abdelsayed, M Wilde, M Sawadogo, and G J Darlington. Autoregulation of the human c/ebp alpha gene by stimulation of upstream stimulatory factor binding. *Mol Cell Biol*, 15(3):1192–202, Mar 1995.

[228] Kelly A. Underwood, James R. Swartz, and Joseph D. Puglisi. Quantitative polysome analysis identifies limitations in bacterial cell-free protein synthesis. *Biotech. Bioeng.*, 91(4):425–35, 2005.

[229] Iván P Uray, Ethan Dmitrovsky, and Powel H Brown. Retinoids and rexinoids in cancer prevention: from laboratory to clinic. *Semin Oncol*, 43(1):49–64, Feb 2016.

[230] Stephen J Van Dien and Mary E Lidstrom. Stoichiometric model for evaluating the metabolic capabilities of the facultative methylotroph methylobacterium extorquens am1, with application to reconstruction of c3 and c4 metabolism. *Biotechnology and bioengineering*, 78(3):296–312, 2002.

[231] Matthew G Vander Heiden, Lewis C Cantley, and Craig B Thompson. Understanding the warburg effect: the metabolic requirements of cell proliferation. *science*, 324(5930):1029–1033, 2009.

[232] C L Varley, E J Bacon, J C Holder, and J Southgate. Foxa1 and irf-1 intermediary transcriptional regulators of ppargamma-induced urothelial cytodifferentiation. *Cell Death Differ*, 16(1):103–14, Jan 2009.

[233] Amit Varma, Brian W Boesch, and Bernhard O Palsson. Biochemical production capabilities of escherichia coli. *Biotechnology and bioengineering*, 42(1):59–73, 1993.

[234] Varnerlab. http://www.varnerlab.org/downloads/.

[235] Michael Vilkhovoy, Nicholas Horvath, Che-hsiao Shih, Joseph Wayman, Kara Calhoun, James Swartz, and Jeffrey Varner. Sequence specific modeling of e. coli cell-free protein synthesis. *bioRxiv*, 2017.

[236] Andreas Von Knethen and Bernhard Brüne. Activation of peroxisome proliferator-activated receptor gamma by nitric oxide in monocytes/macrophages down-regulates p47phox and attenuates the respiratory burst. *J Immunol*, 169(5):2619–26, Sep 2002.

[237] Adina Vultur, Ralf Buettner, Claudia Kowolik, Wei Liang, David Smith, Frank Boschelli, and Richard Jove. Ski-606 (bosutinib), a novel src kinase inhibitor, suppresses migration and invasion of human breast cancer cells. *Molecular cancer therapeutics*, 7(5):1185–1194, 2008.

[238] Aaron S Wallace, Harrison T Supnick, Rodica P Bunaciu, and Andrew Yen. Rrd-251 enhances all-trans retinoic acid (ra)-induced differentiation of hl-60 myeloblastic leukemia cells. *Oncotarget*, 7(29):46401, 2016.

[239] M J Walport. Complement. first of two parts. *N Engl J Med*, 344(14):1058–66, Apr 2001.

[240] M J Walport. Complement. second of two parts. *N Engl J Med*, 344(15):1140–4, Apr 2001.

[241] J. Wang and A. Yen. A novel retinoic acid-responsive element regulates retinoic acid induced BLR1 expression. *Mol. Cell. Biol.*, 24:2423 – 2443, 2004.

[242] J. Wang and A. Yen. A MAPK-positive Feedback Mechanism for BLR1 Signaling Propels Retinoic Acid-triggered Differentiation and Cell Cycle Arrest. *J Biol Chem*, 283:4375–4386, 2008.

[243] Otto Warburg. On the origin of cancer cells. *Science*, 123(3191):309–314, 1956.

[244] R P Warrell, Jr. Retinoid resistance in acute promyelocytic leukemia: new mechanisms, strategies, and implications. *Blood*, 82(7):1949–53, Oct 1993.

[245] RP Jr Warrell. Retinoid resistance in acute promyelocytic leukemia: new mechanisms, strategies, and implications [editorial; comment]. *Blood*, 82(7):1949–1953, 1993.

[246] Justin M Watts and Martin S Tallman. Acute promyelocytic leukemia: what is the new standard of care? *Blood reviews*, 28(5):205–212, 2014.

[247] Joseph A Wayman, Adithya Sagar, and Jeffrey D Varner. Dynamic modeling of cell-free biochemical networks using effective kinetic models. *Processes*, 3(1):138–160, 2015.

[248] Joseph A. Wayman, Adithya Sagar, and Jeffrey D. Varner. Dynamic modeling of cell-free biochemical networks using effective kinetic models. *Processes*, 3(1):138, 2015.

[249] W. Xiong and J.E. Ferrell. A positive-feedback-based bistable 'memory module' that governs a cell fate decision. *Nature*, 426:460–5, 2003.

[250] Edward Yang, Erik van Nimwegen, Mihaela Zavolan, Nikolaus Rajewsky, Mark Schroeder, Marcelo Magnasco, and James E Darnell, Jr. Decay rates of human mrnas: correlation with functional characteristics and sequence attributes. *Genome Res*, 13(8):1863–72, Aug 2003.

[251] T. Yang, Q. Xiong, H. Enslen, R. Davis, and C-W. Chow. Phosphorylation of NFATc4 by p38 mitogen-activated protein kinases. *Mol Cell Biol*, 22:3892–3904, 2002.

[252] A Yen. HL-60 cells as a model of growth and differentiation: the significance of variant cells. *Hematology Review*, 4:5–46, 1990.

[253] A. Yen, S. L. Reece, and K. L. Albright. Dependence of hl-60 myeloid cell differentiation on continuous and split retinoic acid exposures: precommitment memory associated with altered nuclear structure. *J Cell Physiol*, 118(3):277–286, Mar 1984.

[254] A. Yen, M. S. Roberson, S. Varvayanis, and A. T. Lee. Retinoic acid induced mitogen-activated protein (MAP)/extracellular signal-regulated kinase (ERK) kinase-dependent MAP kinase activation needed to elicit HL-60 cell differentiation and growth arrest. *Cancer Res*, 58:3163–3172, 1998.

[255] A. Yen, S. Varvayanis, J. Smith, and T. Lamkin. Retinoic acid induces

expression of SLP-76: expression with c-FMS enhances ERK activation and retinoic acid-induced differentiation/G0 arrest of HL-60 cells. *Eur J Cell Biol*, 85:117–132, 2006.

[256] A Yen, M Williams, JD Platko, C Der, and M Hisaka. Expression of activated raf accelerates cell differentiation and rb protein down-regulation but not hypophosphorylation. *European journal of cell biology*, 65(1):103–113, 1994.

[257] Hiromichi Yuki, Shikiko Ueno, Hiro Tatetsu, Hiroaki Niiro, Tadafumi Iino, Shinya Endo, Yawara Kawano, Yoshihiro Komohara, Motohiro Takeya, Hiroyuki Hata, Seiji Okada, Toshiki Watanabe, Koichi Akashi, Hiroaki Mitsuya, and Yutaka Okuno. Pu.1 is a potent tumor suppressor in classical hodgkin lymphoma cells. *Blood*, 121(6):962–70, Feb 2013.

[258] Nicola Zamboni, Sarah-Maria Fendt, and Uwe Sauer. ^{13}c-based metabolic flux analysis. *Nature Protocols*, 4:878–92, May 2009.

[259] Nehemiah Zewde, Ronald D Gorham Jr, Angel Dorado, and Dimitrios Morikis. Quantitative modeling of the alternative pathway of the complement system. *PloS One*, 11(3):e0152337, 2016.

[260] Yiguo Zhang, Yong-Yeon Cho, Brandon L Petersen, Feng Zhu, and Zigang Dong. Evidence of stat1 phosphorylation modulated by mapks, mek1 and msk1. *Carcinogenesis*, 25(7):1165–1175, Jul 2004.

[261] Ziqing W Zhao, Rahul Roy, J Christof M Gebhardt, David M Suter, Alec R Chapman, and X Sunney Xie. Spatial organization of rna polymerase ii inside a mammalian cell nucleus revealed by reflected light-sheet superresolution microscopy. *Proc Natl Acad Sci U S A*, 111(2):681–6, Jan 2014.

[262] Kai Zhuang, Goutham N Vemuri, and Radhakrishnan Mahadevan. Economics of membrane occupancy and respiro-fermentation. *Molecular systems biology*, 7(1):500, 2011.

[263] Daniel C Zielinski, Neema Jamshidi, Austin J Corbett, Aarash Bordbar, Alex Thomas, and Bernhard O Palsson. Systems biology analysis of drivers underlying hallmarks of cancer cell metabolism. *Scientific reports*, 7:41241, 2017.

[264] S. Zimmermann and K. Moelling. Phosphorylation and regulation of raf by akt (protein kinase b). *Science*, 286(5445):1741–1744, Nov 1999.

[265] Ali R Zomorrodi, Patrick F Suthers, Sridhar Ranganathan, and Costas D Maranas. Mathematical optimization applications in metabolic networks. *Metab Eng*, 14(6):672–86, Nov 2012.